

Лекции по теории формальных языков

Лекция 14.

LR-анализ с неоднозначными грамматиками.
Обработка синтаксических ошибок при LR-анализе.

LR(k)-грамматики.

Итоги

Александр Сергеевич Герасимов

<http://gas-teach.narod.ru>

Кафедра математических и информационных технологий
Санкт-Петербургского академического университета
Российской академии наук.
Весенний семестр 2010/11 учебного года

20 мая 2011 г.

План

- 1 LR-анализ с неоднозначными грамматиками
- 2 Обработка синтаксических ошибок при LR-анализе
- 3 LR(k)-грамматики
- 4 Итоги

План

- 1 LR-анализ с неоднозначными грамматиками
- 2 Обработка синтаксических ошибок при LR-анализе
- 3 LR(k)-грамматики
- 4 Итоги

Пример построения LR(0)-автомата неоднозначной грамматики и разрешения конфликтов

- Расширенная грамматика, полученная из грамматики GA_1 :
(0) $E' \rightarrow E$, (1) $E \rightarrow E + E$, (2) $E \rightarrow E * E$, (3) $E \rightarrow (E)$, (4) $E \rightarrow x$.
- Приоритет $*$ выше приоритета $+$ и эти операции левоассоциативны.
- Построим LR(0)-автомат этой грамматики.
- Попытаемся построить SLR(1)-анализатор.

$FOLLOW(E) = \{+, *,), \dagger\}$.

Конфликты в состояниях:

- ▶ $E^3 = \{[E \rightarrow E + E \cdot], [E \rightarrow E \cdot + E], [E \rightarrow E \cdot * E]\}$,
наверху стека $E^i + E^j$:
 - ★ при очередном входном символе $+$ выполняем свёртку,
 - ★ при очередном входном символе $*$ выполняем перенос;
- ▶ $E^4 = \{[E \rightarrow E * E \cdot], [E \rightarrow E \cdot + E], [E \rightarrow E \cdot * E]\}$,
наверху стека $E^i * E^j$:
 - ★ при очередном входном символе $+$ выполняем свёртку,
 - ★ при очередном входном символе $*$ выполняем свёртку.

Пример построения LR-анализатора для неоднозначной грамматики

(0) $E' \rightarrow E$, (1) $E \rightarrow E + E$, (2) $E \rightarrow E * E$, (3) $E \rightarrow (E)$, (4) $E \rightarrow x$.
 $\text{FOLLOW}(E) = \{+, *,), \vdash\}$.

	ACTION						GOTO
	+	*	x	()	⊢	
							E
E^1	← +	← *				✓	
(← x	← (E^2
x	⊗4	⊗4			⊗4	⊗4	
+			← x	← (E^3
*			← x	← (E^4
E^2	← +	← *			←)		
E^3	⊗1	← *			⊗1	⊗1	
E^4	⊗2	⊗2			⊗2	⊗2	
)	⊗3	⊗3			⊗3	⊗3	
∇			← x	← (E^1

Разрешение конфликта для else из условного оператора

- $\langle \text{оператор} \rangle \rightarrow \text{if } \langle \text{условие} \rangle \text{ then } \langle \text{оператор} \rangle$
 $\langle \text{оператор} \rangle \rightarrow \text{if } \langle \text{условие} \rangle \text{ then } \langle \text{оператор} \rangle \text{ else } \langle \text{оператор} \rangle$
- Спецификация языка программирования Паскаль говорит, что
`if b1 then if b2 then s1 else s2`
эквивалентен
`if b1 then`
 `begin if b2 then s1 else s2 end`
- Конфликт в состоянии LR(0)-автомата, содержащем пункты
 $[\langle \text{оператор} \rangle \rightarrow \text{if } \langle \text{условие} \rangle \text{ then } \langle \text{оператор} \rangle \cdot]$ и
 $[\langle \text{оператор} \rangle \rightarrow \text{if } \langle \text{условие} \rangle \text{ then } \langle \text{оператор} \rangle \cdot \text{else } \langle \text{оператор} \rangle]$,
разрешается в пользу переноса.

План

- 1 LR-анализ с неоднозначными грамматиками
- 2 **Обработка синтаксических ошибок при LR-анализе**
- 3 LR(k)-грамматики
- 4 Итоги

Обнаружение ошибок LR-анализатором

- LR-анализатор обнаруживает ошибку только(?) при обращении к пустой клетке таблицы ACTION.
- Он это делает, как только обработанная часть входа не может быть продолжена до некоторой корректной цепочки, не перенося ошибочный символ в стек.
- В пустую клетку таблицы ACTION можно поместить ссылку на соответствующую процедуру обработки ошибки.
- Обработка ошибок для анализатора арифметических выражений со слайда 5.

Имя ошибки	Сообщение об ошибке	Исправление
ex	отсутствует операнд	добавить x в стек
e+	отсутствует оператор	добавить + в стек
e)	правая скобка без парной ей левой скобки	пропустить) во входной цепочке
e(левая скобка без парной ей правой скобки	добавить) в стек

Таблица LR-анализа с обработкой ошибок

	ACTION						GOTO
	+	*	x	()	⊖	E
E^1	← +	← *	e+	e+	e)	✓	
(ex	ex	← x	← (ex	ex	E^2
x	⊗4	⊗4	⊗4	⊗4	⊗4	⊗4	
+	ex	ex	← x	← (ex	ex	E^3
*	ex	ex	← x	← (ex	ex	E^4
E^2	← +	← *	e+	e+	←)	e(
E^3	⊗1	← *	⊗1	⊗1	⊗1	⊗1	
E^4	⊗2	⊗2	⊗2	⊗2	⊗2	⊗2	
)	⊗3	⊗3	⊗3	⊗3	⊗3	⊗3	
⊖	ex	ex	← x	← (e)	ex	E^1

Протокол работы LR-анализатора с обработкой ошибок

Такт	Содержимое стека	Позиция указателя	Действие
1	∇	$\diamond(x + x))(*x + x \uparrow$	$\leftarrow ($
2	$\nabla($	$(\diamond x + x))(*x + x \uparrow$	$\leftarrow x$
3	$\nabla(x$	$(x \diamond + x))(*x + x \uparrow$	$\otimes 4 E \rightarrow x$
4	$\nabla(E^2$	$(x \diamond + x))(*x + x \uparrow$	$\leftarrow +$
5	$\nabla(E^2 +$	$(x + \diamond x))(*x + x \uparrow$	$\leftarrow x$
6	$\nabla(E^2 + x$	$(x + x \diamond))(*x + x \uparrow$	$\otimes 4 E \rightarrow x$
7	$\nabla(E^2 + E^3$	$(x + x \diamond))(*x + x \uparrow$	$\otimes 1 E \rightarrow E + E$
8	$\nabla(E^2$	$(x + x \diamond))(*x + x \uparrow$	$\leftarrow)$
9	$\nabla(E^2)$	$(x + x) \diamond (*x + x \uparrow$	$\otimes 3 E \rightarrow (E)$
10	∇E^1	$(x + x) \diamond (*x + x \uparrow$	e)
11	∇E^1	$(x + x)) \diamond (*x + x \uparrow$	e+
12	$\nabla E^1 +$	$(x + x)) \diamond (*x + x \uparrow$	$\leftarrow ($
13	$\nabla E^1 + ($	$(x + x))(\diamond * x + x \uparrow$	ex

Протокол работы LR-анализатора с обработкой ошибок: окончание

Такт	Содержимое стека	Позиция указателя	Действие
13	$\nabla E^1 + ($	$(x + x))(\diamond * x + x \dashv$	ex
14	$\nabla E^1 + (x$	$(x + x))(\diamond * x + x \dashv$	$\otimes 4 E \rightarrow x$
15	$\nabla E^1 + (E^2$	$(x + x))(\diamond * x + x \dashv$	$\leftarrow *$
16	$\nabla E^1 + (E^2 *$	$(x + x))(* \diamond x + x \dashv$	$\leftarrow x$
17	$\nabla E^1 + (E^2 * x$	$(x + x))(*x \diamond + x \dashv$	$\otimes 4 E \rightarrow x$
18	$\nabla E^1 + (E^2 * E^4$	$(x + x))(*x \diamond + x \dashv$	$\otimes 2 E \rightarrow E * E$
19	$\nabla E^1 + (E^2$	$(x + x))(*x \diamond + x \dashv$	$\leftarrow +$
20	$\nabla E^1 + (E^2 +$	$(x + x))(*x + \diamond x \dashv$	$\leftarrow x$
21	$\nabla E^1 + (E^2 + x$	$(x + x))(*x + x \diamond \dashv$	$\otimes 4 E \rightarrow x$
22	$\nabla E^1 + (E^2 + E^3$	$(x + x))(*x + x \diamond \dashv$	$\otimes 1 E \rightarrow E + E$
23	$\nabla E^1 + (E^2$	$(x + x))(*x + x \diamond \dashv$	e(
24	$\nabla E^1 + (E^2)$	$(x + x))(*x + x \diamond \dashv$	$\otimes 3 E \rightarrow (E)$
25	$\nabla E^1 + E^3$	$(x + x))(*x + x \diamond \dashv$	$\otimes 1 E \rightarrow E + E$
26	∇E^1	$(x + x))(*x + x \diamond \dashv$	\checkmark (с испр. ошибок.)

План

- 1 LR-анализ с неоднозначными грамматиками
- 2 Обработка синтаксических ошибок при LR-анализе
- 3 LR(k)-грамматики**
- 4 Итоги

Определения LR(k)-грамматики

- Пусть k — целое неотрицательное число, $G = (\Sigma, \Gamma, P, S)$ — КС-грамматика и $G' = (\Sigma, \Gamma', P', S')$ — её расширенная грамматика. G называется *LR(k)-грамматикой*, если из условий

- ▶ $S' \Rightarrow^* \alpha A u \Rightarrow \alpha \beta u$,
- ▶ $S' \Rightarrow^* \gamma B x \Rightarrow \alpha \beta v$ и
- ▶ $\text{FIRST}_k(u) = \text{FIRST}_k(v)$

следует, что $\alpha A v = \gamma B x$ (т. е. $\alpha = \gamma$, $A = B$ и $x = v$).

- Грамматика называется *LR(k)-грамматикой*, если из того, что в выводе r -формы $\alpha \beta u$ последним применяется правило $A \rightarrow \beta$, следует, что это же правило применяется последним в выводе любой r -формы $\alpha \beta v$ такой, что $\text{FIRST}_k(u) = \text{FIRST}_k(v)$.
- Анализатор типа «перенос-свёртка» для LR(k)-грамматики может однозначно определить основу и правило для свёртки по содержимому стека и k очередным символам входной цепочки.
- Грамматика называется *LR-грамматикой*, если она является LR(k)-грамматикой для некоторого целого неотрицательного числа k .

LR(k)-грамматики: факты и примеры

- Класс всех LR(k)-грамматик \subsetneq класс всех LR($k + 1$)-грамматик.
- Грамматика $G_1 = \{S \rightarrow Ba^k b \mid Ca^k c, B \rightarrow d, C \rightarrow d\}$ является LR($k + 1$)-грамматикой, но не является LR(k)-грамматикой.
 - ▶ $S \Rightarrow^* Ba^k b \Rightarrow \underbrace{\varepsilon}_{\alpha} \underbrace{d}_{\beta} \underbrace{a^k b}_{u}$,
 - ▶ $S \Rightarrow^* Ca^k c \Rightarrow \underbrace{\varepsilon}_{\alpha} \underbrace{d}_{\beta} \underbrace{a^k c}_{v}$, $\text{FIRST}_k(u) = \text{FIRST}_k(v)$.
- Грамматика $G_2 = \{S \rightarrow Bb \mid Cc, B \rightarrow Ba \mid \varepsilon, C \rightarrow Ca \mid \varepsilon\}$ однозначна, но не является LR-грамматикой.
 - ▶ $S \Rightarrow^* Ba^k b \Rightarrow \underbrace{\varepsilon}_{\alpha} \underbrace{\varepsilon}_{\beta} \underbrace{a^k b}_{u}$,
 - ▶ $S \Rightarrow^* Ca^k c \Rightarrow \underbrace{\varepsilon}_{\alpha} \underbrace{\varepsilon}_{\beta} \underbrace{a^k c}_{v}$, $\text{FIRST}_k(u) = \text{FIRST}_k(v)$.

Теорема

Класс всех LL(k)-грамматик является собственным подклассом класса всех LR(k)-грамматик.

Теорема

Любой LR-язык является $SLR(1)$ -языком.

Следствие

Класс LR-языков совпадает с каждым из классов $SLR(1)$ -, $LALR(1)$ -, $LR(k)$ -языков, где $k \geq 1$.

Теорема

Класс LR-языков совпадает с классом языков, распознаваемых ДМПА.

План

- 1 LR-анализ с неоднозначными грамматиками
- 2 Обработка синтаксических ошибок при LR-анализе
- 3 LR(k)-грамматики
- 4 **Итоги**

Иерархия Хомского.

Соотношения между классами языков

Граматики (типа 0)

- контекстно-зависимые грамматики
- — неукорачивающие КС-грамматики
- КС-грамматики
- — праволинейные грамматики

Класс праволинейных языков =

класс распознаваемых ДКА языков =

класс распознаваемых НКА языков =

класс распознаваемых ε -НКА языков =

класс определяемых регулярными выражениями языков \subsetneq

класс распознаваемых ДМПА языков \subsetneq

класс распознаваемых НМПА языков = класс КС-языков \subsetneq

класс КЗ-языков \subsetneq класс разрешимых языков \subsetneq

класс языков типа 0 = класс перечислимых языков.

Иерархия некоторых классов грамматик,
часто используемых для синтаксического анализа.
Соотношения между классами языков

КС-грамматики

— однозначные грамматики

— — LR(1)-грамматики

— — — LL(1)-грамматики

— — — LALR(1)-грамматики

— — — — SLR(1)-грамматики

— — — — — LR(0)-грамматики

— — — — — СП-грамматики

— — — — — — ПП-грамматики

Класс LL(1)-языков \subsetneq класс LL-языков \subsetneq

класс распознаваемых ДМПА языков =

класс LR-языков = класс LR(1)-языков = класс LALR(1)-языков =

класс SLR(1)-языков \supsetneq класс СП-языков = класс ПП-языков.

Литература

Основная литература

- Замятин А. П., Шур А. М. Языки, грамматики, распознаватели: Учебное пособие. Екатеринбург : Изд-во Урал. ун-та, 2007 (электронный вариант книги — на <http://elar.usu.ru>, поиск).

Дополнительная литература

- Ахо А., Лам М., Сети Р., Ульман Дж. Компиляторы: принципы, технологии и инструментарий. М.: ООО "И.Д. Вильямс", 2008.
- Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. М.: Мир, 1978.
- Мартыненко Б. К. Языки и трансляции: Учеб. пособие. СПб.: Издательство С.-Петербургского университета, 2004 (электронный вариант книги — на <http://www.math.spbu.ru/user/mbk>).