

Gramatyki SLR, LR(1) i LALR

Języki formalne i techniki translacji - Wykład 11

Maciek Gębala

18 grudnia 2018

Maciek Gębala Gramatyki SLR, LR(1) i LALR

Tablice analizatorów SLR

- Sytuacją $LR(0)$ nazywamy produkcję z gramatyki G z kropką w jakimś miejscu prawej strony.
- Z produkcji $A \rightarrow XYZ$ możemy otrzymać cztery sytuacje: $A \rightarrow \cdot XYZ$, $A \rightarrow X \cdot YZ$, $A \rightarrow XY \cdot Z$ i $A \rightarrow XYZ \cdot$, a z produkcji $A \rightarrow \varepsilon$ jedną $A \rightarrow \cdot$.
- Każdą sytuację możemy reprezentować parą liczb: numer produkcji i pozycja kropki.
- Konstruujemy z gramatyki deterministyczny automat skończony rozpoznający odpowiednie prefiksy.
- Sytuacje można traktować jak stany automatu niedeterministycznego.
- Rozważane gramatyki uzupełniamy o specjalną produkcję początkową $S' \rightarrow S$.

Maciek Gębala Gramatyki SLR, LR(1) i LALR

Operacja domknięcia

Jeśli I jest zbiorem sytuacji z gramatyki G to domknięcie(I) jest zbiorem sytuacji otrzymanych z I przy zastosowaniu reguł

- 1 Każda z sytuacji z I należy do domknięcia(I).
- 2 Jeśli $A \rightarrow \alpha \cdot B\beta$ jest w domknięciu(I), a $B \rightarrow \gamma$ jest produkcją to do domknięcia(I) dodajemy $B \rightarrow \cdot \gamma$. Powtarzamy, dopóki można dodać nowe elementy.

Maciek Gębala Gramatyki SLR, LR(1) i LALR

Przykład

$$\begin{aligned} E' &\rightarrow E \\ E &\rightarrow E + T \mid T \\ T &\rightarrow T * F \mid F \\ F &\rightarrow (E) \mid id \end{aligned}$$

- Niech $I = \{[E' \rightarrow \cdot E]\}$.
- Wtedy domknięcie(I) zawiera

$$\begin{aligned} E' &\rightarrow \cdot E \\ E &\rightarrow \cdot E + T \mid \cdot T \\ T &\rightarrow \cdot T * F \mid \cdot F \\ F &\rightarrow \cdot (E) \mid \cdot id \end{aligned}$$

Maciek Gębala Gramatyki SLR, LR(1) i LALR

Notatki

Notatki

Notatki

Notatki

Jeśli I jest zbiorem sytuacji a X symbolem z gramatyki, to przejście(I, X) jest domknięciem zbioru wszystkich sytuacji $A \rightarrow \alpha X \cdot \beta$ takich, że $A \rightarrow \alpha \cdot X \beta$ jest w I .

Przykład

$$\begin{aligned} E' &\rightarrow E \\ E &\rightarrow E + T \mid T \\ T &\rightarrow T * F \mid F \\ F &\rightarrow (E) \mid id \end{aligned}$$

- Niech $I = \{[E' \rightarrow E \cdot], [E \rightarrow E \cdot + T]\}$.
- Wtedy przejście($I, +$) zawiera

$$\begin{aligned} E &\rightarrow E + \cdot T \\ T &\rightarrow \cdot T * F \mid \cdot F \\ F &\rightarrow \cdot (E) \mid \cdot id \end{aligned}$$

Konstruowanie zbiorów sytuacji

- $C \leftarrow \{ \text{domknięcie}(\{[S' \rightarrow \cdot S]\}) \}$
- Dopóki C może się powiększyć to dla każdego zbioru sytuacji $I \in C$ i każdego symbolu X takiego, że przejście(I, X) $\neq \emptyset$ dodaj przejście(I, X) do C .

Przykład

- $l_0 : E' \rightarrow \cdot E, E \rightarrow \cdot E + T, E \rightarrow \cdot T, T \rightarrow \cdot T * F, T \rightarrow \cdot F, F \rightarrow \cdot (E), F \rightarrow \cdot id$
- $l_1 : E' \rightarrow E \cdot, E \rightarrow E \cdot + T$
- $l_2 : E \rightarrow T \cdot, T \rightarrow T \cdot * F$
- $l_3 : T \rightarrow F \cdot$
- $l_4 : F \rightarrow (\cdot E), E \rightarrow \cdot E + T, E \rightarrow \cdot T, T \rightarrow \cdot T * F, T \rightarrow \cdot F, F \rightarrow \cdot (E), F \rightarrow \cdot id$
- $l_5 : F \rightarrow id \cdot$
- $l_6 : E \rightarrow E + \cdot T, T \rightarrow \cdot T * F, T \rightarrow \cdot F, F \rightarrow \cdot (E), F \rightarrow \cdot id$
- $l_7 : T \rightarrow T * \cdot F, F \rightarrow \cdot (E), F \rightarrow \cdot id$
- $l_8 : F \rightarrow (E \cdot), E \rightarrow E \cdot + T$
- $l_9 : E \rightarrow E + T \cdot, T \rightarrow T \cdot * F$
- $l_{10} : T \rightarrow T * F \cdot$
- $l_{11} : F \rightarrow (E) \cdot$

Funkcja przejścia dla utworzonego DFA

	E	T	F	$($	$)$	$+$	$*$	id
l_0	l_1	l_2	l_3	l_4				l_5
l_1						l_6		
l_2							l_7	
l_3								
l_4	l_8	l_2	l_3					l_5
l_5								
l_6		l_9	l_3	l_4				l_5
l_7			l_{10}	l_4				l_5
l_8					l_{11}	l_6		
l_9							l_7	
l_{10}								
l_{11}								

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Konstrukcja tablicy analizatora SLR

- 1 Zbuduj $C = \{l_0, \dots, l_n\}$ - rodzinę zbiorów sytuacji dla G' .
- 2 Stan i budujemy z l_i . Akcje analizatora dla stanu i wyznaczamy następująco:
 - 1 Jeśli $[A \rightarrow \alpha \cdot a\beta] \in l_i$, a jest terminalem i przejście(l_i, a) = l_j to akcja[i, a] = j .
 - 2 Jeśli $[A \rightarrow \alpha \cdot] \in l_i$ to akcja[i, a] = $A \rightarrow \alpha$ dla wszystkich $a \in FOLLOW(A)$ ($A \neq S'$).
 - 3 Jeśli $[S' \rightarrow S \cdot] \in l_i$ to akcja[$i, \$$] = ACC .
- 3 Jeśli przejście(l_i, A) = l_j to przejście[i, A] = j .
- 4 Stan startowy analizatora to stan odpowiadający zbiorowi sytuacji zawierającemu $[S' \rightarrow \cdot S]$.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Przykład

$S \rightarrow L = R \mid R, L \rightarrow *R \mid id, R \rightarrow L$

$l_0 : S' \rightarrow \cdot S, S \rightarrow \cdot L = R, S \rightarrow \cdot R, L \rightarrow \cdot * R, L \rightarrow \cdot id, R \rightarrow \cdot L; l_1 : S' \rightarrow S \cdot; l_2 : S \rightarrow L \cdot = R, R \rightarrow L \cdot; l_3 : S \rightarrow R \cdot; l_4 : L \rightarrow * \cdot R, R \rightarrow \cdot L, L \rightarrow * \cdot R, L \rightarrow \cdot id; l_5 : L \rightarrow id \cdot; l_6 : S \rightarrow L = \cdot R, R \rightarrow \cdot L, L \rightarrow * \cdot R, L \rightarrow \cdot id; l_7 : L \rightarrow *R \cdot; l_8 : R \rightarrow L \cdot; l_9 : S \rightarrow L = R \cdot;$

Co to jest akcja[2, =]?

- akcja[2, =] = 6 bo przejście($l_2, =$) = l_6
- akcja[2, =] = $R \rightarrow L$ bo $= \in FOLLOW(R)$
- **Konflikt redukcja/przesunięcie**

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Kanoniczne tablice analizatorów LR

- Przechowujemy w stanie więcej informacji aby wykluczyć niektóre konflikty.
- Rozszerzamy definicję sytuacji przez dodanie do produkcji z kropką terminala lub $\$$. (Sytuacja $LR(1)$)
- Drugą składową nazywamy podglądem sytuacji.
- Redukcję $A \rightarrow \alpha$ wykonujemy tylko dla tych symboli wejściowych a dla których sytuacja $[A \rightarrow \alpha \cdot, a]$ wyznaczyła stan z wierzchołka stosu.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Notatki

Notatki

Notatki

Konstruowanie sytuacji $LR(1)$

domknięcie(I)

Dopóki można dodać nowy element do I dla każdej sytuacji $[A \rightarrow \alpha \cdot B\beta, a] \in I$, każdej produkcji $B \rightarrow \gamma \in G'$ i każdego terminala $b \in FIRST(\beta a)$ dodaj $[B \rightarrow \cdot \gamma, b]$ do I .

przejście(I, X)

J – zbiór sytuacji $[A \rightarrow \alpha X \cdot \beta, a]$ takich, że $[A \rightarrow \alpha \cdot X\beta, a] \in I$.
 $przejście(I, X) = domknięcie(J)$.

Konstruowanie zbiorów sytuacji $LR(1)$

- 1 $C \leftarrow \{ domknięcie(\{[S' \rightarrow \cdot S, \$]\}) \}$
- 2 Dopóki C może się powiększyć to dla każdego zbioru sytuacji $I \in C$ i każdego symbolu X takiego, że $przejście(I, X) \neq \emptyset$ dodaj $przejście(I, X)$ do C .

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Konstrukcja kanonicznej tablicy analizatora LR

- 1 Zbuduj $C = \{I_0, \dots, I_n\}$ – rodzinę zbiorów sytuacji $LR(1)$ dla G' .
- 2 Stan i budujemy z I_i . Akcje analizatora dla stanu i wyznaczamy następująco:
 - 1 Jeśli $[A \rightarrow \alpha \cdot a\beta, b] \in I_i$, a jest terminalem i $przejście(I_i, a) = I_j$ to akcja $[i, a] = j$.
 - 2 Jeśli $[A \rightarrow \alpha \cdot, b] \in I_i$ to akcja $[i, b] = A \rightarrow \alpha$ ($A \neq S'$).
 - 3 Jeśli $[S' \rightarrow S \cdot, \$] \in I_i$ to akcja $[i, \$] = ACC$.
- 3 Jeśli $przejście(I_i, A) = I_j$ to $przejście[i, A] = j$.
- 4 Stan startowy analizatora to stan odpowiadający zbiorowi sytuacji zawierającemu $[S' \rightarrow \cdot S, \$]$.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Gramatyki $LALR$

- Tablice analizatorów $LR(1)$ są często bardzo duże – tablice $LALR$ są przeważnie mniejsze.
- Główna idea to łączenie takich stanów $LR(1)$ które nie powodują konfliktów.

Jądro sytuacji

Jądrem sytuacji nazywamy te sytuacje których prawa strona nie zaczyna się od kropki oraz sytuację $S' \rightarrow \cdot S$. (Rozpatrujemy sytuację $LR(0)$).

Operacja domknięcia zbioru sytuacji $LR(0)$ dodaje tylko sytuacje spoza jądra.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Przykład

$S' \rightarrow S \quad S \rightarrow CC \quad C \rightarrow cC|d$

Zbiory sytuacji

$I_0 : [S' \rightarrow \cdot S, \$], [S \rightarrow \cdot CC, \$], [C \rightarrow \cdot cC, c/d], [C \rightarrow \cdot d, c/d];$
 $I_1 : [S' \rightarrow S \cdot, \$];$
 $I_2 : [S \rightarrow C \cdot C, \$], [C \rightarrow \cdot cC, \$], [C \rightarrow \cdot d, \$];$
 $I_3 : [C \rightarrow c \cdot C, c/d], [C \rightarrow \cdot cC, c/d], [C \rightarrow \cdot d, c/d];$
 $I_4 : [C \rightarrow d \cdot, c/d];$
 $I_5 : [S \rightarrow CC \cdot, \$];$
 $I_6 : [C \rightarrow c \cdot C, \$], [C \rightarrow \cdot cC, \$], [C \rightarrow \cdot d, \$];$
 $I_7 : [C \rightarrow d \cdot, \$];$
 $I_8 : [C \rightarrow cC \cdot, c/d];$
 $I_9 : [C \rightarrow cC \cdot, \$];$

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Przykład

Kanoniczna tablica analizatora

stan	akcja			przejście	
	<i>c</i>	<i>d</i>	<i>\$</i>	<i>S</i>	<i>C</i>
<i>s</i> ₀	<i>s</i> ₃	<i>s</i> ₄		<i>s</i> ₁	<i>s</i> ₂
<i>s</i> ₁			<i>acc</i>		
<i>s</i> ₂	<i>s</i> ₆	<i>s</i> ₇			<i>s</i> ₅
<i>s</i> ₃	<i>s</i> ₃	<i>s</i> ₄			<i>s</i> ₈
<i>s</i> ₄	<i>r</i> ₃	<i>r</i> ₃			
<i>s</i> ₅			<i>r</i> ₁		
<i>s</i> ₆	<i>s</i> ₆	<i>s</i> ₇			<i>s</i> ₉
<i>s</i> ₇			<i>r</i> ₃		
<i>s</i> ₈	<i>r</i> ₂	<i>r</i> ₂			
<i>s</i> ₉			<i>r</i> ₂		

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Przykład

Porównajmy stany s_3 i s_6 , s_4 i s_7 , oraz s_8 i s_9 . Poszczególne pary mają te same jądra sytuacji (różnią się one tylko podglądanym symbolem).

Czy połączenie stanów może spowodować konflikt?

Tak. Jest to jednak mało prawdopodobne dla błędu przesunięcie/redukcja. Może natomiast zajść błąd redukcja/redukcja. Stąd nie wszystkie gramatyki $LR(1)$ są gramatykami $LALR(1)$.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Przykład powstania konfliktu

$S' \rightarrow S$, $S \rightarrow aAd|bBd|aBe|bAe$, $A \rightarrow c$, $B \rightarrow c$

Tworząc zbiory sytuacji otrzymamy m.in. $\{[A \rightarrow c, d], [B \rightarrow c, e]\}$ i $\{[A \rightarrow c, e], [B \rightarrow c, d]\}$. Żaden z tych zbiorów nie wywołuje konfliktów.

Ich suma $\{[A \rightarrow c, d/e], [B \rightarrow c, d/e]\}$ wywołuje konflikt.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Metoda budowania tablic LALR

- 1 Zbuduj $C = \{I_0, \dots, I_n\}$ – rodzinę zbiorów sytuacji $LR(1)$.
- 2 Dla każdego jądra zbiorów sytuacji znajdź wszystkie zbiory o tym samym jądrze i zastąp je przez ich sumę.
- 3 Niech $C' = \{J_0, \dots, J_m\}$. Utwórz na podstawie tego zbioru tablicę akcji i przejść. Jeśli nastąpił konflikt – przerwij (uznaj, że gramatyka nie jest typu $LALR(1)$).

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Przykład

Tablica analizatora $LALR(1)$

stan	akcja			przejście	
	c	d	$\$$	S	C
s_0	s_{36}	s_{47}		s_1	s_2
s_1			acc		
s_2	s_{36}	s_{47}			s_5
s_{36}	s_{36}	s_{47}			s_{89}
s_{47}	r_3	r_3	r_3		
s_5			r_1		
s_{89}	r_2	r_2	r_2		

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Różnice w działaniu $LR(1)$ i $LALR(1)$

Dla poprawnych danych obie metody działają identycznie dając to samo wyprowadzenie.

Dla niepoprawnych wejść analiza $LALR(1)$ może wykonać pewne redukcje po zgłoszeniu błędu przez analizę $LR(1)$. Ale analiza $LALR(1)$ nie wykona już żadnego przesunięcia.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Przykład

$S' \rightarrow S \quad S \rightarrow CC \quad C \rightarrow cC|d$

Słowo: ccd

$LR(1)$ odłoży na stosie $s_0 c s_{36} c s_{36} d s_4$ i widząc $\$$ zgłosi błąd.

- $LALR(1)$ odłoży na stosie $s_0 c s_{36} c s_{36} d s_{47}$ i widząc $\$$ dokona redukcji $C \rightarrow d$ i zmieni stos na $s_0 c s_{36} c s_{36} C s_{89}$.
- Teraz wykona redukcję $C \rightarrow cC$ i otrzyma $s_0 c s_{36} C s_{89}$.
- Następna redukcja $C \rightarrow cC$ i otrzyma $s_0 C s_2$.
- I dopiero teraz wykryje błąd.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Obsługa błędów w analizie LR

Kanoniczny analizator LR przed zgłoszeniem błędu nigdy nie wykona żadnej redukcji. Analizatory SLR i LALR mogą wykonać pewną liczbę redukcji ale nie przesuną na stos błędnego symbolu wejściowego.

Tryb paniki

Przeglądamy stos w dół aż napotkamy stan s z przejściem dla pewnego nieterminala A . Następnie wyrzucamy z wejścia symbole aż znajdziemy symbol a który legalnie może występować po A . Odkładamy na stos przejście $[s, A]$ i kontynuujemy analizę.

Uzupełnienie tablicy analizatora o obsługę błędów

W puste miejsca tablicy wstawiamy odpowiednie akcje korygujące stos i wejście.

Maciek Gębala

Gramatyki SLR, LR(1) i LALR

Notatki

Notatki

Notatki

Notatki

- Rola analizatora składniowego.
- Deterministyczne gramatyki bezkontekstowe.
- Analiza zstępująca: gramatyki $LL(k)$.
- Analiza wstępująca: gramatyki operatorowe, SLR, LR i LALR.
- Obsługa błędów.

Notatki

Notatki

Notatki

Notatki
