Pontificia Universidad Católica de Chile
Departamento de Ciencia de la Computación
IIC2223 – Teoría de Autómatas y Lenguajes Formales
Segundo semestre de 2024
Profesor: Cristian Riveros
Ayudante: Amaranta Salas

Ayudantia 14

First, Follow y LL(k)

Problema 1

1. Considere la gramática

$$\mathcal{G} = \big(\{S\}, \{a,b\}, \{S \;\to\; SS \mid aSb \mid bSa \mid \epsilon \;\}, S\big)$$

   Para cada variable X de $\mathcal{G}$, calcule $\texttt{first}_1(X)$ y $\texttt{follow}_1(X)$ usando los algoritmos vistos en clases.

2. Considere la gramática

$$\mathcal{G} = \left(\{S, M, P\}, \{a, (,), *\}, \begin{Bmatrix} S & \to & P \mid M \mid a \\ P & \to & (S) \\ M & \to & S * S \end{Bmatrix}, S\right)$$

   Para cada variable X de $\mathcal{G}$, calcule $\texttt{first}_1(X)$ usando los algoritmos vistos en clases.

Solución

1. Construimos $\texttt{first}_1^i(S)$ iterativamente:

$$\begin{aligned}
\texttt{first}_1^0(S) &= \{\varepsilon\} \\
\texttt{first}_1^1(S) &= \{\varepsilon, \\
&\qquad a \odot_1 \texttt{first}_1^0(S) \odot_1 b, \\
&\qquad b \odot_1 \texttt{first}_1^0(S) \odot_1 a, \\
&\qquad \texttt{first}_1^0(S) \odot_1 \texttt{first}_1^0(S)\} = \{\varepsilon, a, b\} \\
\texttt{first}_1^2(S) &= \{\varepsilon, a, b, \\
&\qquad a \odot_1 \texttt{first}_1^1(S) \odot_1 b, \\
&\qquad b \odot_1 \texttt{first}_1^1(S) \odot_1 a, \\
&\qquad \texttt{first}_1^1(S) \odot_1 \texttt{first}_1^1(S)\} = \{\varepsilon, a, b\}
\end{aligned}$$

   Como $\texttt{first}_1^2(S) = \texttt{first}_1^1(S)$, entonces $\texttt{first}_1(S) = \texttt{first}_1^1(S) = \{\varepsilon, a, b\}$. Ahora calculamos el follow:

$$\begin{aligned}
\texttt{follow}_1^0(S) &= \{\#, a, b\} \\
\texttt{follow}_1^1(S) &= \{\texttt{first}_1(b) \odot_1 \texttt{follow}_1^0(S), && S \to aSb \\
&\qquad \texttt{first}_1(a) \odot_1 \texttt{follow}_1^0(S), && S \to bSa \\
&\qquad \texttt{first}_1(S) \odot_1 \texttt{follow}_1^0(S), && S \to SS \\
&\qquad \} = \{b, a, \#\}
\end{aligned}$$

2. Construimos $\texttt{first}_1^i(X)$ iterativamente para cada $X \in \mathcal{G}$:

$$
\begin{aligned}
\texttt{first}_1^0(S) &= \{a\} \\
\texttt{first}_1^1(S) &= \{a\} \\
\texttt{first}_1^1(P) &= \{(\odot_1 \texttt{first}_1^0(S) \odot_1)\} = \{(\} \\
\texttt{first}_1^1(M) &= \{\texttt{first}_1^0(S) \odot_1 * \odot_1 \texttt{first}_1^0(S)\} = \{a\} \\
\texttt{first}_1^2(S) &= \{\texttt{first}_1^1(P), \\
&\qquad \texttt{first}_1^1(M), \\
&\qquad a\} = \{a, (\} \\
\texttt{first}_1^2(M) &= \{\texttt{first}_1^1(S) \odot_1 * \odot_1 \texttt{first}_1^1(S)\} = \{a, (\} \\
\texttt{first}_1^2(P) &= \{(\odot_1 \texttt{first}_1^1(S) \odot_1)\} = \{(\} \\
&\ldots
\end{aligned}
$$

Problema 2

Una gramática libre de contexto $\mathcal{G}$ se dice **unambigua** si para toda palabra $w \in \mathcal{L}(\mathcal{G})$ existe exactamente una derivación por la izquierda de $\mathcal{G}$ sobre w, en otras palabras, existe un **único** árbol de derivación para w.

Demuestre que para toda gramática libre de contexto $\mathcal{G}$, si $\mathcal{G}$ es $LL(k)$ para algún k, entonces $\mathcal{G}$ es unambigua.

Solución

Demostraremos por contrapositivo, es decir, que si $\mathcal{G}$ no es unambigua, entonces $\mathcal{G}$ no es LL(k) para algún $k > 0$. Sea $\mathcal{G}$ una gramática no unambigua. Por contradicción, suponga que $\mathcal{G}$ es LL(k) para algún $k > 0$.

Como $\mathcal{G}$ es una gramática NO unambigua, entonces existe $w \in \mathcal{L}(\mathcal{G})$ tal que existen 2 o más derivaciones de w según $\mathcal{G}$. Suponga que estas son:

- $S \underset{lm}{\Longrightarrow} \alpha_1 \underset{lm}{\Longrightarrow} \alpha_2 \underset{lm}{\Longrightarrow} \ldots \underset{lm}{\Longrightarrow} \alpha_n \underset{lm}{\Longrightarrow} w$
- $S \underset{lm}{\Longrightarrow} \beta_1 \underset{lm}{\Longrightarrow} \beta_2 \underset{lm}{\Longrightarrow} \ldots \underset{lm}{\Longrightarrow} \beta_n \underset{lm}{\Longrightarrow} w$

Sea i el menor natural tal que $\alpha_i \neq \beta_i$. Entonces,

1. $S \overset{*}{\underset{lm}{\Longrightarrow}} uX\gamma \underset{lm}{\Longrightarrow} u\gamma_1\gamma \overset{*}{\underset{lm}{\Longrightarrow}} w$, con $uX\gamma = \alpha_{i-1}$ y $u\gamma_1\gamma = \alpha_i$.
2. $S \overset{*}{\underset{lm}{\Longrightarrow}} uX\gamma \underset{lm}{\Longrightarrow} u\gamma_2\gamma \overset{*}{\underset{lm}{\Longrightarrow}} w$, con $uX\gamma = \beta_{i-1}$ y $u\gamma_2\gamma = \beta_i$.

Como $\alpha_i \neq \beta_i$, entonces $\gamma_1 \neq \gamma_2$ $(*)$.

Luego, sea v tal que $w = uv$. Si consideramos $v = v_1 = v_2$, como $\mathcal{G}$ se supuso LL(k), se cumple el ítem 1, el ítem 2 y $v_1|_k = v_2|_k$, por definición se deduce que $\gamma_1 = \gamma_2$. Pero, esto contradice $(*)$.

Por lo tanto, si $\mathcal{G}$ no es unambigua, entonces $\mathcal{G}$ no puede ser LL(k).

Problema 3

1. Demuestre que existe una gramática libre de contexto $\mathcal{G}$ en forma normal de Chomsky y $k \geq 3$, tales que $\mathcal{G}$ es $LL(k)$, pero no $LL(k)$ fuerte.
2. Demuestre que para toda gramática libre de contexto $\mathcal{G}$ en forma normal de Chomsky y $k \leq 2$, si $\mathcal{G}$ es $LL(k)$ entonces es $LL(k)$ fuerte.

Solución

1. Para demostrar que existe una CFG en CNF que es $LL(k)$, pero no $LL(k)$ fuerte para algún $k \geq 3$, bastará con encontrar una CFG que cumpla la condición buscada.

   Podemos encontrar la gramática basándonos en el ejemplo visto en clases para una gramática cualquiera:

$$\begin{aligned} S &\to aXaa \mid bXba \\ X &\to b \mid \varepsilon \end{aligned}$$

   Es claro que esta gramática no está en CNF, pero podemos seguir la misma idea formando la siguiente gramática $\mathcal{G}$:

$$\begin{aligned} S &\to AY_1 \mid BY_2 \\ Y_1 &\to XZ_1 \\ Y_2 &\to XZ_2 \\ X &\to CB \mid C \\ Z_1 &\to AA \\ Z_2 &\to BA \\ A &\to a \\ B &\to b \\ C &\to c \end{aligned}$$

   Luego, podemos seguir la misma idea de clases, pero para $LL(3)$ encontrando que:

   - $S \overset{*}{\Rightarrow} aXZ_1,\ X \Rightarrow CB \wedge X \Rightarrow C$:

$$\texttt{first}_3(CBZ_1) \cap \texttt{first}_3(CZ_1) = \{cba\} \cap \{caa\} = \varnothing$$

   - $S \overset{*}{\Rightarrow} aXZ_2,\ X \Rightarrow CB \wedge X \Rightarrow C$:

$$\texttt{first}_3(CBZ_2) \cap \texttt{first}_3(CZ_2) = \{cbb\} \cap \{cba\} = \varnothing$$

   Por lo tanto, $\mathcal{G}$ es $LL(3)$.

   Analizando ahora la gramática para ver si es $LL(3)$ fuerte, con $S \overset{*}{\Rightarrow} aXZ_1,\ X \Rightarrow CB \wedge X \Rightarrow C$ podemos ver lo siguiente:

$$\begin{aligned} \texttt{first}_3(CB) \odot_3 \texttt{follow}_3(X) \cap \texttt{first}_3(C) \odot_3 \texttt{follow}_3(X) &= \\ \texttt{first}_3(CB) \odot_3 (\texttt{first}_3(Z_1) \cup \texttt{first}_3(Z_2)) \cap \texttt{first}_3(C) \odot_3 (\texttt{first}_3(Z_1) \cup \texttt{first}_3(Z_2)) &= \\ \{cb\} \odot_3 \{aa, ba\} \cap \{c\} \odot_3 \{aa, ba\} &= \\ \{cba, cbb\} \cap \{caa, cba\} &= \\ \{cba\} \end{aligned}$$

   Por lo tanto, $\mathcal{G}$ no es $LL(3)$ fuerte.

2. Para demostrar que toda gramática en CNF que es $LL(k)$ es también $LL(k)$ fuerte para $k \leq 2$, bastará demostrar esta condición para $k = 2$, pues el caso $k = 1$ se demostró en clases.

   Sea $\mathcal{G}$ una gramática $LL(2)$ en CNF. Sabemos que cumplirá que para todas dos reglas distintas $Y \to \gamma_1$, $Y \to \gamma_2$ y para todo $S \overset{*}{\Rightarrow} uY\beta$ se tiene que:

$$\texttt{first}_2(\gamma_1\beta) \cap \texttt{first}_2(\gamma_2\beta) = \varnothing$$

   Podemos demostrar por casos que esta gramática será $LL(2)$ fuerte.

(a) Suponga que $\texttt{first}_2(\gamma_1) \cap \Sigma^{\leq 1} \neq \varnothing \;\; \wedge \;\; \texttt{first}_2(\gamma_2) \cap \Sigma^{\leq 1} \neq \varnothing$. Para que algún elemento de $\texttt{first}_2(\gamma_1)$ tenga largo 1, tiene que cumplirse que $\gamma_1 = a$, con $a \in \Sigma$, ya que sabemos que $\mathcal{G}$ está en CNF.

Análogamente, tendremos que $\gamma_2 = b$, con $b \in \Sigma$ y como $Y \to \gamma_1$ y $Y \to \gamma_2$ deben ser reglas distintas, es claro que $a \neq b$ y, por tanto:

$$\texttt{first}_2(\gamma_1) \odot_2 \texttt{follow}_2(Y) \cap \texttt{first}_2(\gamma_2) \odot_2 \texttt{follow}_2(Y) = \varnothing$$

(b) Suponga que $\texttt{first}_2(\gamma_1) \cap \Sigma^{\leq 1} = \varnothing \;\; \wedge \;\; \texttt{first}_2(\gamma_2) \cap \Sigma^{\leq 1} = \varnothing$. Como todos los elementos de $\texttt{first}_2(\gamma_1)$ y $\texttt{first}_2(\gamma_2)$ tienen largo 2, tendremos que:

$$\begin{aligned}\texttt{first}_2(\gamma_1) \odot_2 \texttt{follow}_2(Y) &= \texttt{first}_2(\gamma_1) \\ &= \texttt{first}_2(\gamma_1\beta)\end{aligned}$$

Ánalogamente, $\texttt{first}_2(\gamma_2) \odot_2 \texttt{follow}_2(Y) = \texttt{first}_2(\gamma_2\beta)$. Entonces se cumple que:

$$\texttt{first}_2(\gamma_1) \odot_2 \texttt{follow}_2(Y) \cap \texttt{first}_2(\gamma_2) \odot_2 \texttt{follow}_2(Y) = \varnothing$$

(c) Suponga que $\texttt{first}_2(\gamma_1) \cap \Sigma^{\leq 1} \neq \varnothing \;\; \wedge \;\; \texttt{first}_2(\gamma_2) \cap \Sigma^{\leq 1} = \varnothing$. Al igual que para el primer caso, tendremos que $\gamma_1 = a$, con $a \in \Sigma$ y, al igual que en el segundo caso, tendremos que $\texttt{first}_2(\gamma_2) = \texttt{first}_2(\gamma_2\beta) = \texttt{first}_2(\gamma_2\beta')$ para todo $\beta, \beta' \in (V \cup \Sigma)^*$.

Luego, como $\mathcal{G}$ es $LL(2)$, tenemos que para todo $\beta' \in (V \cup \Sigma)^*$:

$$\begin{aligned}\varnothing &= \texttt{first}_2(\gamma_1\beta) \cap \texttt{first}_2(\gamma_2\beta) \\ &= \texttt{first}_2(\gamma_1\beta) \cap \texttt{first}_2(\gamma_2) \\ &= \texttt{first}_2(\gamma_1\beta) \cap \texttt{first}_2(\gamma_2\beta')\end{aligned}$$

Lo que significa que:

$$\begin{aligned}&\texttt{first}_2(\gamma_1) \odot_2 \texttt{follow}_2(Y) \cap \texttt{first}_2(\gamma_2) \odot_2 \texttt{follow}_2(Y) \\ &= \bigcup_{S \overset{*}{\underset{\texttt{lm}}{\Longrightarrow}} uY\beta} \texttt{first}_2(\gamma_1\beta) \cap \bigcup_{S \overset{*}{\underset{\texttt{lm}}{\Longrightarrow}} uY\beta'} \texttt{first}_2(\gamma_1\beta') \\ &= \varnothing\end{aligned}$$

Finalmente, como para todos los casos se cumple que $\texttt{first}_2(\gamma_1) \odot_2 \texttt{follow}_2(Y) \cap \texttt{first}_2(\gamma_2) \odot_2 \texttt{follow}_2(Y) = \varnothing$, si una gramática $\mathcal{G}$ en CNF es $LL(k)$, también será $LL(k)$ fuerte para todo $k \leq 2$.