6.2 y 6.3 Formas normales y Membresía

José de Jesús Lavalle Martínez

Benemérita Universidad Autónoma de Puebla
Facultad de Ciencias de la Computación
Lenguajes Formales y Autómatas CCOS 014

Contenido

Formas normales

En el contexto de Lenguajes Formales y Autómatas, una **forma normal** es aquella que restringe la forma de las producciones de las gramáticas sin alterar su poder generativo.

Formas normales

En el contexto de Lenguajes Formales y Autómatas, una **forma normal** es aquella que restringe la forma de las producciones de las gramáticas sin alterar su poder generativo.

Se espera que esa restricción sea de utilidad teórica y/o práctica para, por ejemplo, establecer propiedades sobre las gramáticas que estamos estudiando.

Concretamente, la forma normal de Chomsky tiene la restricción de que el lado derecho de una regla tenga a lo más dos símbolos, de la siguiente manera.

Definición 1

Una gramática libre de contexto **está en forma normal de Chomsky** si todas sus producciones son de la forma

$$A \to BC$$

o

$$A \to a,$$

donde $A, B, C \in V$ y $a \in T$.

Ejemplo forma normal de Chomsky

Ejemplo 2

La gramática dada por las producciones

$$S \to AS|a,$$
$$A \to SA|b$$

está en forma normal de Chomsky. La gramática dada por

$$S \to AS|AAS,$$
$$A \to SA|aa$$

no está en forma normal de Chomsky, las producciones $S \to AAS$ y $A \to aa$ violan las condiciones de la definición 1.

Teorema forma normal de Chomsky I

Teorema 3

Cualquier gramática libre de contexto $G = (V, T, S, P)$ *con* $\lambda \notin L(G)$ *tiene una gramática equivalente* $\widehat{G} = (\widehat{V}, \widehat{T}, S, \widehat{P})$ *en forma normal de Chomsky.*

Demostración: Dado el teorema 17 de la sesión pasada, podemos asumir sin pérdida de generalidad que G no tiene producciones-λ, ni producciones unitarias, ni símbolos inútiles. Haremos la construcción de $\widehat{G}$ en dos pasos.

Paso 1: Construya una gramática $G_1 = (V_1, T, S, P_1)$ a partir de G considerando todas las producciones en P de la forma

$$A \to x_1 x_2 \cdots x_n \tag{1}$$

donde cada x_i es un símbolo en V o en T. Si $n = 1$, entonces x_1 debe ser un terminal ya que G no tiene producciones unitarias. En este caso, ponga la producción en P_1.

Teorema forma normal de Chomsky II

Si $n \geq 2$, introduzca nuevas variables B_a para cada $a \in T$. Para cada producción de P en la forma (1) ponga en P_1 la producción

$$A \to C_1 C_2 \cdots C_n$$

donde

$$C_i = x_i, \text{ si } x_i \in V,$$

y

$$C_i = B_a, \text{ si } x_i = a.$$

Para todo B_a también ponga en P_1 la producción

$$B_a \to a.$$

Teorema forma normal de Chomsky III

Esta parte del algoritmo quita todos los terminales de las producciones cuyos lados derechos tienen longitud mayor a uno, reemplazándolos con variables nuevas. Al final de este paso tenemos una gramática G_1 donde todas sus producciones tienen la forma

$$A \to a, \tag{2}$$

o

$$A \to C_1 C_2 \cdots C_n, \tag{3}$$

donde $C_i \in V_1$.

Como consecuencia del Teorema 4 de la sesión pasada (Regla de sustitución) tenemos que

$$L(G_1) = L(G)$$

Teorema forma normal de Chomsky IV

Paso 2: Se introducen variables adicionales para reducir la longitud de los lados derechos de las producciones, donde sea necesario. Primero, todas las producciones de la forma (2) así como las de la forma (3) con $n = 2$ se ponen dentro de $\widehat{P}$.

Para cada producción de la forma (3) con $n > 2$, se introducen variables $D_1, D_2, \cdots$ y se ponen en $\widehat{P}$ las producciones

$$\begin{aligned} A &\to C_1 D_1 \\ D_1 &\to C_2 D_2 \\ &\vdots \\ D_{n-2} &\to C_{n-1} C_n. \end{aligned}$$

La gramática resultante está en forma normal de Chomsky. Aplicando varias veces el Teorema 4 de la sesión pasada (Regla de sustitución) tenemos que

$$L(\widehat{G}) = L(G).$$

Ejemplo forma normal de Chomsky V

Ejemplo 4

Convierta la grámatica, con las producciones siguientes, a forma normal de Chomsky.

$$
\begin{aligned}
S &\to ABa,\\
A &\to aab,\\
B &\to Ac.
\end{aligned}
$$

Ejemplo forma normal de Chomsky V

Ejemplo 4

Convierta la grámatica, con las producciones siguientes, a forma normal de Chomsky.

$$S \to ABa,$$
$$A \to aab,$$
$$B \to Ac.$$

Solución: Como se requiere, la gramática no tiene producciones-λ ni producciones unitarias.

Ejemplo 4

Convierta la grámatica, con las producciones siguientes, a forma normal de Chomsky.

$$S \to ABa,$$
$$A \to aab,$$
$$B \to Ac.$$

Solución: Como se requiere, la gramática no tiene producciones-λ ni producciones unitarias.
En el primer paso introducimos las variables nuevas B_a, B_b y B_c, y hacemos las sustituciones adecuadas para obtener

Ejemplo forma normal de Chomsky V

Ejemplo 4

Convierta la grámatica, con las producciones siguientes, a forma normal de Chomsky.

$$
\begin{aligned}
S &\rightarrow ABa, \\
A &\rightarrow aab, \\
B &\rightarrow Ac.
\end{aligned}
$$

Solución: Como se requiere, la gramática no tiene producciones-λ ni producciones unitarias.
En el primer paso introducimos las variables nuevas B_a, B_b y B_c, y hacemos las sustituciones adecuadas para obtener

$$
\begin{aligned}
S &\rightarrow ABB_a, & B_a &\rightarrow a, \\
A &\rightarrow B_aB_aB_b, & B_b &\rightarrow b, \\
B &\rightarrow AB_c, & B_c &\rightarrow c.
\end{aligned}
$$

Ejemplo forma normal de Chomsky VI

$$
\begin{aligned}
S &\to ABB_a, & B_a &\to a,\\
A &\to B_aB_aB_b, & B_b &\to b,\\
B &\to AB_c, & B_c &\to c.
\end{aligned}
$$

En el segundo paso, se introducen variables adicionales para que las dos primeras producciones queden en forma normal de Chomsky y así obtener el resultado final.

$$
\begin{aligned}
S &\to AD_1, & B &\to AB_c,\\
D_1 &\to BB_a, & B_a &\to a,\\
A &\to B_aD_2, & B_b &\to b,\\
D_2 &\to B_aB_b, & B_c &\to c.
\end{aligned}
$$

Forma normal de Greibach I

La forma normal de Greibach no restringe la longitud de los lados derechos de las producciones, sino la posición en la que pueden aparecer los símbolos terminales y las variables.

Definición 5

Una gramática libre de contexto se dice que **está en forma normal de Greibach** si todas las producciones tienen la forma

$$A \to ax,$$

donde $a \in T$ y $x \in V^*$.

Forma normal de Greibach II

Ejemplo 6

La gramática

$$
\begin{aligned}
S &\to AB,\\
A &\to aA|bB|b,\\
B &\to b
\end{aligned}
$$

no está en forma normal de Greibach.

Forma normal de Greibach II

Ejemplo 6

La gramática

$$
\begin{aligned}
S &\to AB,\\
A &\to aA|bB|b,\\
B &\to b
\end{aligned}
$$

no está en forma normal de Greibach.

Pero aplicando las sustituciones del Teorema 4 de la sesión anterior, obtenemos la gramática equivalente

$$
\begin{aligned}
S &\to aAB|bBB|bB,\\
A &\to aA|bB|b,\\
B &\to b
\end{aligned}
$$

Ejemplo 7

Convierta la gramática

$$S \rightarrow abSb|aa$$

a su forma normal de Greibach.

Introducimos nuevas variables A y B como sinónimos para los terminales a y b, por supuesto también agregamos las producciones correspondientes, Sustituyendo los terminales por sus respectivas variables obtenemos la gramática

$$\begin{aligned} S &\rightarrow aBSB|aA, \\ A &\rightarrow a, \\ B &\rightarrow b \end{aligned}$$

que está en forma normal de Greibach.

El algoritmo CYK I

- El algoritmo toma su nombre por las iniciales de los apellidos de sus inventores J. Cocke, D. H. Younger y T. Kasami.

El algoritmo CYK I

- El algoritmo trabaja sólo si la gramática está en forma normal de Chomsky.

El algoritmo CYK II

Asuma que tenemos una gramática libre de contexto $G = (V, T, S, P)$ en forma normal de Chomsky y una cadena

$$w = a_1 a_2 \cdots a_n.$$

Defina subcadenas

$$w_{ij} = a_i \cdots a_j,$$

y subconjuntos de V

$$V_{ij} = \left\{ A \in V : A \stackrel{*}{\Rightarrow} w_{ij} \right\}$$

así, $w \in L(G)$ si y sólo si $S \in V_{1n}$.

El algoritmo CYK III

Para calcular V_{ij}, observe que $A \in V_{ii}$ si y sólo si G contiene una producción $A \to a_i$. Por lo tanto, V_{ii} se puede calcular para todo $1 \leq i \leq n$ inspeccionando a w y las producciones de la gramática.

Para continuar, note que para todo $j > i$, A deriva w_{ij} si y sólo si existe una producción $A \to BC$, con $B \stackrel{*}{\Rightarrow} w_{ik}$ y $C \stackrel{*}{\Rightarrow} w_{k+1j}$ para algún k con $i \leq k, k < j$. En otras palabras

$$V_{ij} = \bigcup_{k \in \{i, i+1, \cdots, j-1\}} \{A : A \to BC, \text{ con } B \in V_{ik}, C \in V_{k+1j}\} \quad (4)$$

El algoritmo CYK IV

Una inspección a los índices en (4) muestra que se pueden usar para calcular todos los V_{ij} si procedemos en la secuencia.

1. Calcule $V_{11}, V_{22}, \cdots, V_{nn}$,
2. Calcule $V_{12}, V_{23}, \cdots, V_{n-1n}$,
3. Calcule $V_{13}, V_{24}, \cdots, V_{n-2n}$,
4. Calcule $V_{14}, V_{25}, \cdots, V_{n-3n}$,
5. $\vdots$
6. Calcule $V_{1n-2}, V_{2n-1}, V_{3n}$,
7. Calcule V_{1n-1}, V_{2n},
8. Calcule V_{1n}.

Ejemplo algoritmo CYK I

Ejemplo 8

Determine si la cadena $w = aabbb$ está en el lenguaje generado por la gramática dada por las producciones

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

Solución: Ya que $w_{11} = a$ el conjunto V_{11} de todas las variables que derivan inmediatamente a a es $V_{11} = \{A\}$, así obtenemos que

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

.

Ejemplo algoritmo CYK II

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

Ahora usamos (4) para obtener

$$V_{12} = \{A : A \to BC, B \in V_{11}, C \in V_{22}\},$$

la única variable de V_{11} es A y la única variable de V_{22} también es A, por lo tanto buscamos AA como parte derecha de alguna de las reglas, como no está entonces $V_{12} = \emptyset$.

Ejemplo algoritmo CYK III

$$S \rightarrow AB,$$
$$A \rightarrow BB|a,$$
$$B \rightarrow AB|b.$$

Ejemplo algoritmo CYK III

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

Para V_{23}, buscamos AB del lado derecho de las producciones, vemos que tanto S como B derivan AB, por lo tanto $V_{23} = \{S, B\}$.

Ejemplo algoritmo CYK III

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

Para V_{34}, buscamos BB del lado derecho de las producciones, vemos que sólo A deriva a BB, por lo tanto $V_{34} = \{A\}$.

Ejemplo algoritmo CYK III

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

Para V_{45}, buscamos BB del lado derecho de las producciones, vemos que sólo A deriva a BB, por lo tanto $V_{45} = \{A\}$.

Ejemplo algoritmo CYK IV

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

$$V_{12} = \emptyset, V_{23} = \{S, B\}, V_{34} = \{A\}, V_{45} = \{A\}.$$

Para V_{13} buscamos del lado derecho de las producciones a AS y AB, la primera no se encuentra pero a la segunda la pueden derivar S y B, así $V_{13} = \{S, B\}$. Siguiendo de esta manera obtenemos:

$$V_{13} = \{S, B\}, V_{24} = \{A\}, V_{35} = \{S, B\}.$$

Ejemplo algoritmo CYK V

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

$$V_{12} = \emptyset, V_{23} = \{S, B\}, V_{34} = \{A\}, V_{45} = \{A\}.$$

$$V_{13} = \{S, B\}, V_{24} = \{A\}, V_{35} = \{S, B\}.$$

Para V_{14} se tienen que analizar $V_{11}V_{24}, V_{12}V_{34}, V_{13}V_{44}$.
Para V_{25} se tienen que analizar $V_{22}V_{35}, V_{23}V_{45}, V_{24}V_{55}$.
Para V_{15} se tienen que analizar $V_{11}V_{25}, V_{12}V_{35}, V_{13}V_{45}, V_{14}V_{55}$.

Ejemplo algoritmo CYK VI

$$S \to AB,$$
$$A \to BB|a,$$
$$B \to AB|b.$$

$$V_{11} = \{A\}, V_{22} = \{A\}, V_{33} = \{B\}, V_{44} = \{B\}, V_{55} = \{B\}.$$

$$V_{12} = \emptyset, V_{23} = \{S, B\}, V_{34} = \{A\}, V_{45} = \{A\}.$$

$$V_{13} = \{S, B\}, V_{24} = \{A\}, V_{35} = \{S, B\}.$$

$$V_{14} = \{A\}, V_{25} = \{S, B\}.$$

$$V_{15} = \{S, B\}.$$

Como $S \in V_{15}$ entonces $aabbb \in L(G)$.

Teorema 9

El algoritmo CYK tiene orden de complejidad $O(n^3)$.

Demostración: Note que el número de conjuntos V_{ij} que se tienen que calcular es exactamente $n(n+1)/2$ donde n es la longitud de la cadena w. Tambíen note que a lo más n términos requieren ser analizados en (4). Por lo tanto CYK está en $O(n^3)$.

Ejercicios I

1. Convierta las siguientes gramáticas a forma normal de Chomsky
   a. $S \rightarrow aSS|a|b.$
   b. $S \rightarrow aSb|Sab|ab.$
   c. $$S \rightarrow aSaaA|A,$$ $$A \rightarrow abA|bb.$$
   d. $$S \rightarrow baAB,$$ $$A \rightarrow bAB|\lambda$$ $$B \rightarrow BAa|A|\lambda.$$

Ejercicios II

e.

$$S \to AB|aB,$$
$$A \to abb|\lambda$$
$$B \to bbA.$$

2. Convierta las siguientes gramáticas a forma normal de Greibach

   a.

$$S \to aSb|bSa|a|b|ab$$

   b.

$$S \to aSb|ab|bb$$

   c.

$$S \to ab|aS|aaS|aSS$$

d.

$$
\begin{aligned}
S &\to ABb|a|b, \\
A &\to aaA|B, \\
B &\to bAb.
\end{aligned}
$$

3. Use el algoritmo CYK para saber si las siguientes palabras están en el lenguaje generado por la gramática del ejercicio 8 de esta presentación.
   1. $aaabbb$.
   2. $aaaabbb$