

Autómatas y Lenguajes Formales 2016-1

Maestría en Ciencia e Ingeniería de la Computación UNAM

Tema 7: Lenguajes no regulares, el lema del bombeo y el teorema de Myhill-Nerode

Dr. Favio Ezequiel Miranda Perea

`favio@ciencias.unam.mx`

Facultad de Ciencias UNAM

29 de enero de 2020



¿ Cuántos lenguajes regulares hay?

$$REG = \{L \subseteq \Sigma^* \mid L \text{ es regular} \}$$

- Dado que un lenguaje L es un subconjunto de Σ^* , existen tantos lenguajes como elementos en $\mathcal{P}(\Sigma^*)$.
- Puesto que Σ^* es infinito numerable, es decir, es del tamaño del conjunto \mathbb{N} de los números naturales, entonces $\mathcal{P}(\Sigma^*)$ es del tamaño del conjunto de los números reales \mathbb{R} .
- Por otra parte, existen sólo tantos lenguajes regulares como números naturales, $|REG| = |\mathbb{N}|$.
- De manera que el conjunto $\mathcal{P}(\Sigma^*) - REG$ no puede ser numerable, pues unión de numerables sigue siendo numerable.
- Es decir, hay tantos lenguajes no regulares como números reales.



¿ Cuántos lenguajes regulares hay?

$$REG = \{L \subseteq \Sigma^* \mid L \text{ es regular} \}$$

- Existen sólo tantos lenguajes regulares como números naturales:
 - ▶ La idea de la prueba es enumerar lexicográficamente todos los AFD posibles con alfabeto de entrada Σ , es decir, primero los autómatas con un sólo estado, luego los de dos estados, etc,etc..
 - ▶ Esto implica que el número de lenguajes regulares es a lo más numerable.
 - ▶ Además claramente es numerable pues hay una infinidad numerable de lenguajes regulares, por ejemplo

$$\{a\}, \{aa\}, \{aaa\}, \dots$$



El lema del bombeo

Propiedad de lenguajes regulares

Lema (Lema del Bombeo)

Si L es un lenguaje regular infinito entonces existe un número $n \in \mathbb{N}$, llamado constante de bombeo para L , tal que para cualquier cadena w de L con $|w| \geq n$ existen cadenas u, v, x tales que:

- 1 $w = uvx$
- 2 $v \neq \varepsilon$
- 3 $|uv| \leq n$
- 4 $\forall i \geq 0, uv^i x \in L.$

Demostración.

En clase.



Pruebas de no regularidad

Lema del bombeo

- Para probar que un lenguaje L no es regular se procede por contradicción usando del lema del bombeo como sigue:
 - ▶ Si L fuera regular existiría una constante de bombeo n .
 - ▶ Cualquier palabra $w \in L$ se descompone como $w = uvx$, $v \neq \varepsilon$, $|uv| \leq n$.
 - ▶ Se llega a una contradicción como sigue:
 - ★ Por el lema del bombeo $uv^i x \in L$ para toda $i \geq 0$.
 - ★ Por la definición particular de L , existe alguna i tal que $uv^i x \notin L$.
- Debemos observar que encontrar la i adecuada depende del problema particular y no hay un método general, pero generalmente basta con valores pequeños de i .



$L = \{a^i b^i \mid i \in \mathbb{N}\}$ no es regular

Lema del bombeo

- Supóngase que L es regular y sea n una constante de bombeo.
- Considérese la palabra $w = a^n b^n$ y su descomposición $w = uvx$ con $v \neq \varepsilon$, $|uv| \leq n$.
- Entonces u, v constan de puras a 's, digamos $u = a^k$, $v = a^\ell$, $k \geq 0, \ell \geq 1$.
- De manera que $x = a^{n-k-\ell} b^n$.
- Si hacemos $i = 2$ tenemos que $uv^2x \in L$ por el lema del bombeo.
- Pero, por otra parte

$$uv^2x = a^k a^\ell a^\ell a^{n-k-\ell} b^n = a^{n+\ell} b^n \notin L$$



$L = \{a^{i^2} \mid i \in \mathbb{N}\}$ no es regular

Lema del bombeo

- Supongase que L es regular y sea n una constante de bombeo.
- Considérese la palabra $w = a^{n^2}$ y su descomposición $w = uvx$ con $v \neq \varepsilon$, $|uv| \leq n$.
- Digamos que $u = a^k$, $v = a^\ell$, $1 \leq \ell \leq n$.
- De manera que $x = a^{n^2 - k - \ell}$.



$L = \{a^{i^2} \mid i \in \mathbb{N}\}$ no es regular

Lema del bombeo

- Si hacemos $i = 2$ tenemos que $uv^2x \in L$ por el lema del bombeo.
- Pero, por otra parte

$$uv^2x = a^k a^\ell a^\ell a^{n^2-k-\ell} = a^{n^2+\ell}$$

- Y observemos que $n^2 < n^2 + \ell < (n+1)^2$ pues $1 \leq \ell$ y

$$\ell \leq n < 2n+1 = (n+1)^2 - n^2$$

- De manera que $n^2 + \ell$ no es de la forma m^2 , por lo que $a^{n^2+\ell} \notin L$



$L = \{w \in \{a, b\}^* \mid w = w^R\}$ no es regular

Lema del bombeo

- Supongase que L es regular y sea n una constante de bombeo.
- Considérese la palabra $w = a^n b^n a^n$ y su descomposición $w = uvx$ con $v \neq \varepsilon$, $|uv| \leq n$.
- Entonces u, v constan de puras a 's, digamos $u = a^k$, $v = a^\ell$, $\ell \geq 1$.
- De manera que $x = a^{n-k-\ell} b^n a^n$.
- Si hacemos $i = 2$ tenemos que $uv^2x \in L$ por el lema del bombeo.
- Pero, por otra parte

$$uv^2x = a^k a^\ell a^\ell a^{n-k-\ell} b^n a^n = a^{n+\ell} b^n a^n \notin L$$



Relaciones de indistinguibilidad entre cadenas

Teorema de Myhill-Nerode

- Considerense las siguientes relaciones de equivalencia sobre Σ^* relacionadas a un lenguaje dado L y a un autómata finito determinista dado M .

- ▶ $x \equiv_L y$ si y sólo si

$$\forall z \in \Sigma^* (xz \in L \Leftrightarrow yz \in L)$$

- ▶ $x \equiv_M y$ si y sólo si

$$\delta^*(q_0, x) = \delta^*(q_0, y)$$

- Si $x \equiv_M y$ entonces se dice que x, y son cadenas indistinguibles según M .
- Si $x \equiv_L y$ entonces se dice que x, y son cadenas indistinguibles para L .



Relación \equiv_L

Ejemplos para $L = \{a^n b^n \mid n \in \mathbb{N}\}$

$x \equiv_L y$ si y sólo si $\forall z \in \Sigma^* (xz \in L \Leftrightarrow yz \in L)$

- $a^4 b^3 \equiv_L a^3 b^2$ pues

$$\forall z \in \Sigma^* (a^4 b^3 z \in L \Leftrightarrow z = b \Leftrightarrow a^3 b^2 z \in L)$$

- $a^2 b^2 \not\equiv_L a^3 b^2$ pues para $z = \varepsilon$ se tiene

$$a^2 b^2 z \in L \text{ y } a^3 b^2 z \notin L$$

- $a^4 b^2 \not\equiv_L a^3 b^2$ pues para $z = b$, se tiene

$$a^4 b^2 z \notin L \text{ y } a^3 b^2 z \in L$$

- $abb \equiv_L baba$



Relación \equiv_L

Ejemplos

- Sea $L = \{w \in \{a, b\}^* \mid w \text{ tiene un número par de } a\}$
 - ▶ $x \equiv_L y$ si y sólo si x y y tienen la misma paridad.
 - ▶ Por lo tanto existen dos clases de equivalencia: $[\varepsilon] = L$ y $[a] = \{a, b\}^* - L$
- Sea $L = \{w \in \{0, 1\}^* \mid w \text{ empieza y termina con el mismo símbolo}\}$
 - ▶ $x \equiv_L y$ si y sólo si x y y comienzan con un mismo símbolo y terminan con un mismo símbolo
 - ▶ $x \equiv_L y$ si y sólo si $x = awb$ y $y = avb$ con $a, b \in \{0, 1\}$
 - ▶ Se sigue que \equiv_L tiene 5 clases de equivalencia:

$$[\varepsilon] = \varepsilon \quad [0] = 0 + 0(0 + 1)^*0 \quad [1] = 1 + 1(0 + 1)^*1$$

$$[01] = 0(0 + 1)^*1 \quad [10] = 1(0 + 1)^*0$$



Relación \equiv_L

Ejemplos

- Sea $L = \{a^n b^n \mid n \in \mathbb{N}\}$
- La relación \equiv_L tiene una infinidad de clases de equivalencia, por ejemplo:

$$[\varepsilon], [a], [a^2], \dots, [a^n], \dots$$

- Todas estas clases son diferentes pues si $i \neq j$ entonces $a^i \not\equiv_L a^j$
- Esto es claro pues si $z = b^j$ entonces $a^i z \in L$ pero $a^j z \notin L$.



Relaciones de indistinguibilidad entre cadenas

Teorema de Myhill-Nerode

- Por lo general no hay relación alguna entre L y M . La relación \equiv_L puede definirse para cualquier lenguaje L aún cuando este no sea regular.
- Sin embargo, en el caso particular en que $L = L(M)$ se cumple que \equiv_M es un refinamiento de \equiv_L , es decir

$$\forall x, y \in \Sigma^* (x \equiv_M y \rightarrow x \equiv_L y).$$

- Esta proposición nos deja ver la más importante limitación de los autómatas finitos, el hecho de que carecen de memoria más allá de lo que recuerde el estado actual.
- Si $x \equiv_M y$ entonces por la proposición anterior $x \equiv_{L(M)} y$, por lo que ninguna cadena w procesada después de x o y permitirá que M determine cuál de x o y se procesó anteriormente.



Invariancia de las relaciones \equiv_M, \equiv_L

Teorema de Myhill-Nerode

- Una relación de equivalencia \equiv sobre Σ^* es invariante por la derecha si y sólo si

$$\forall x, y, w \in \Sigma^* (x \equiv y \rightarrow xw \equiv yw).$$

- La relación \equiv_L es invariante por la derecha.
- Lema de continuación: Sean $x, y \in \Sigma^*$. Si $\delta^*(q_0, x) = \delta^*(q_0, y)$ entonces para cualquier $z \in \Sigma^*$, se cumple que

$$\delta^*(q_0, xz) = \delta^*(q_0, yz)$$

- Del lema anterior se sigue que la relación \equiv_M es invariante por la derecha.



Propiedades de la relación \equiv_M

- Recordemos que el índice de una relación de equivalencia \equiv es el número de clases de equivalencia generadas por \equiv .
- Dado un AFD $M = \langle Q, \Sigma, q_0, \delta, F \rangle$ se cumple lo siguiente:
 - ▶ La relación \equiv_M es invariante por la derecha.
 - ▶ La relación \equiv_M es de índice finito.
 - ▶ $L(M)$ es la unión de algunas de las clases de equivalencia de la relación \equiv_M .



El Teorema de Myhill-Nerode

Propiedad de lenguajes regulares

Teorema (Myhill-Nerode)

Sea $L \subseteq \Sigma^*$. Las siguientes condiciones son equivalentes:

- 1 L es regular.
- 2 Existe una relación de equivalencia \equiv sobre Σ^* , invariante por la derecha y de índice finito, tal que L es la unión de algunas de las clases de equivalencia de \equiv .
- 3 La relación de equivalencia \equiv_L tiene índice finito.



Lema del índice finito

Teorema de Myhill-Nerode

- Por el teorema de Myhill-Nerode para mostrar que un lenguaje L no es regular basta mostrar que L no es de índice finito.
- Es decir, basta con ver que \equiv_L tiene una infinidad de clases de equivalencia.
- Esto se hace explícito mediante el siguiente lema que es una consecuencia directa del teorema de Myhill-Nerode.
- Lema del índice finito: Sea $L \subseteq \Sigma^*$ un lenguaje regular infinito. Cualquier conjunto $S \subseteq \Sigma^*$ suficientemente grande contiene al menos dos cadenas distintas, $x, y \in S$ tales que $x \equiv_L y$.



Conjuntos estafadores¹

Pruebas de no regularidad

- Las pruebas de no regularidad se sirven de la contrapositiva del lema del índice finito.
- Hallando un conjunto S que no cumpla la propiedad del lema, habremos probado que L no es regular.
- Un conjunto infinito $S \subseteq \Sigma^*$ es un conjunto estafador para L si y sólo si para cualesquiera $x, y \in S$ existe una cadena $z \in \Sigma^*$ tal que una y sólo una de xz y yz pertenecen a L .
- Es decir, S es un conjunto estafador para L si y sólo si

$$\forall x, y \in S (x \not\equiv_L y).$$

- Para mostrar que un lenguaje L no es regular basta construir un conjunto estafador para L .

¹En inglés *fooling set*



$L = \{a^i b^i \mid i \in \mathbb{N}\}$ no es regular

Teorema de Myhill-Nerode

- Basta hallar un conjunto estafador para $L = \{a^n b^n \mid n \in \mathbb{N}\}$.
- Sea $S = \{a^k \mid k \in \mathbb{N}\}$, veamos que S es un conjunto estafador.
- Si $a^i, a^j \in S$ con $i \neq j$ entonces claramente

$$a^i b^i \in L \text{ y } a^j b^i \notin L.$$

- Por lo tanto

$$a^i \not\equiv_L a^j$$

y así S es un conjunto estafador para L .



$L = \{a^{i^2} \mid i \in \mathbb{N}\}$ no es regular

Teorema de Myhill-Nerode

- Basta hallar un conjunto estafador para $L = \{a^{i^2} \mid i \in \mathbb{N}\}$
- Sea $S = L$, veamos que S es un conjunto estafador.
- Sean $a^{i^2}, a^{j^2} \in S$ con $j > i$.
 - ▶ Por un lado tenemos que $a^{i^2} a^{2i+1} = a^{i^2+2i+1} = a^{(i+1)^2} \in L$
 - ▶ Por otra parte, $a^{j^2} a^{2i+1} = a^{j^2+2i+1} \notin L$ puesto que

$$j^2 < j^2 + 2i + 1 < j^2 + 2j + 1 = (j+1)^2.$$

- Por lo tanto

$$a^{i^2} \not\equiv_L a^{j^2}$$

y S es un conjunto estafador para L .

