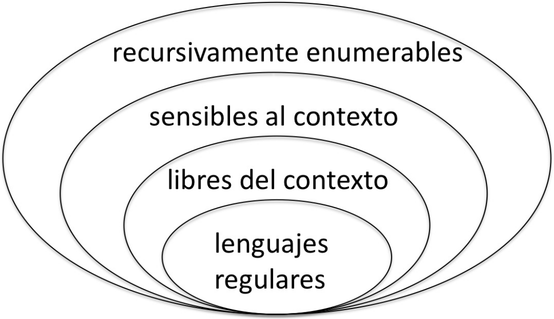
Teoría de Lenguajes

Lema de Pumping y propiedades de lenguajes regulares

DC-UBA

1do Cuatrimestre 2024

Jerarquía de Chomsky



Tipo	Lenguaje	Autómata	Normas de producción de gramáticas	Ejemplos
0	recursivamente enumerable (LRE)	Máquina de Turing	$\alpha A \beta \to \delta$	$L = \{w \mid w$ describe una máquina de Turing$\}$
1	dependiente del contexto (LSC)	Autómata linealmente acotado	$\alpha A \beta \to \alpha \gamma \beta$	$L = \{a^n b^n c^n \mid n > 0\}$
2	independiente del contexto (LLC)	Autómata con pila	$A \to \gamma$	$L = \{a^n b^n \mid n > 0\}$
3	regular (LR)	Autómata finito	$A \to \mathrm{a}$ y $A \to \mathrm{a}B$	$L = \{a^n \mid n \geq 0\}$

Lenguajes regulares

En dónde estamos parados

- Formalismos para especificar lenguajes regulares
  - Autómatas finitos (AFD, AFND, AFND-λ)
  - Expresiones regulares (ER)
  - Gramáticas regulares
- Equivalencias entre los formalismos
  - AFD $\iff$ AFND-λ
  - ER $\iff$ AF
- Propiedades de los lenguajes regulares

Lenguajes regulares
Propiedades

Sean L_1 y L_2 lenguajes regulares, L_1 definido sobre el alfabeto Σ. Los siguientes también son lenguajes regulares:

- Unión: $L_1 \cup L_2$
- Concatenación: $L_1 L_2$
- Intersección: $L_1 \cap L_2$
- Complemento: ${L_1}^c = \Sigma^* \setminus L_1$
- Reversa: ${L_1}^r$

Lenguajes regulares

Más propiedades

Sean $n \in \mathbb{N}$ y $L_1, ..., L_n$ lenguajes regulares, sea L_{nr} un lenguaje no-regular

- Unión finita: $\bigcup_{i=1}^{n} L_i$ es regular
- Intersección finita: $\bigcap_{i=1}^{n} L_i$ es regular
- Unión disjunta: $L_1 \overset{d}{\cup} L_{nr}$ es no-regular

No existe propiedad para los siguientes casos:

- Unión infinita de lenguajes regulares
- Intersección infinita de lenguajes regulares
- Unión de lenguajes no-regulares
  - $L_{nr} \cup L_{nr}$ es no-regular
  - $L_{nr} \cup L_{nr}^{c}$ es regular
- Intersección de lenguajes no-regulares

Lema de pumping

Sabemos probar que un lenguaje es regular.
Usamos el lema de pumping para probar que un lenguaje NO es regular.

Enunciado formal

L regular $\rightarrow$

$$\exists p > 0\ (\forall \alpha\ ((\alpha \in L \wedge |\alpha| \geq p) \rightarrow$$
$$\exists x \exists y \exists z\ (\alpha = xyz \wedge |xy| \leq p \wedge |y| \geq 1 \wedge \forall i \geq 0\ (xy^{\mathrm{i}}z \in L))))$$

En castellano

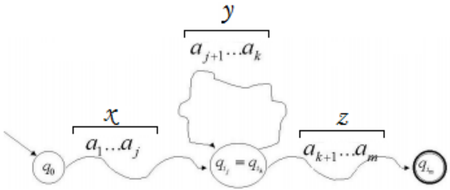
Sea L un lenguaje regular. Existe entonces una longitud mínima p tal que, todas las cadenas $\alpha \in L$ que superan o igualan dicha longitud, pueden ser escritas de la forma $\alpha = xyz$ donde

$$|xy| \leq p \qquad |y| \geq 1 \qquad \forall i \geq 0, xy^{\mathrm{i}}z \in L$$

Lema de pumping

Un poco de teoría

- Al ser L regular, existe al menos un AFD de estados mínimos que lo reconoce. Tomemos p el número de estados de este autómata
- Todas las cadenas $\alpha \in L$ cuya longitud supera al tamaño del autómata (p) pasan por algún estado al menos dos veces (al menos un ciclo)
- Existe una descomposición xyz donde x es la parte de la cadena reconocida hasta el 1er estado que se repite, y es la parte de la cadena desde ese estado hasta la siguente vez que toca el estado, z es el resto



- Para cualquier repetición de y, la cadena resultante ($xz, xyz, xyyz$) va a continuar perteneciendo a L (*pumping / bombeo*)

Lema de pumping

Usos y no usos

Abuso de notación: llamaremos ”*pumping*” al consecuente del lema.

Si L es regular $\Rightarrow$ L cumple ”*pumping*”
Si L NO cumple ”*pumping*” $\Rightarrow$ L NO es regular (por contrarrecíproco)

Entonces si probamos que un lenguaje no cumple ”*pumping*”, estamos dando una demostración válida de que no es regular.
Vamos a usar el lema con esta intención.

La vuelta (L cumple ”*pumping*” $\Rightarrow$ L es regular) NO es cierta (ver ejercicio 2 de la práctica 5).

Lema de pumping

Traduciendo y llegando a la técnica

L no cumple "*pumping*" equivale a probar que la sig. fórmula es verdadera

$$\forall p > 0 \; (\exists \alpha \; (\alpha \in L \wedge |\alpha| \geq p \wedge \\ \forall x \forall y \forall z \; (\alpha = xyz \wedge |xy| \leq p \wedge |y| \geq 1 \wedge \exists i \geq 0 \; (xy^{i}z \notin L))))$$

Para todo $p > 0$

Existe α tal que α pertenece al lenguaje L, $|\alpha| \geq p$

Para toda descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$

Existe un $i \geq 0$ tal que $xy^{i}z \notin L$

1. Me dan un $p > 0$
2. Elijo una cadena α perteneciente a L con longitud mayor o igual a p
3. Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$
4. Elijo un i mayor o igual a cero tal que $xy^{i}z$ no pertenezca a L

Lema de pumping

Atención, concentración

1. Me dan un $p > 0$
2. Elijo una cadena α perteneciente a L con longitud mayor o igual a p
3. Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$
4. Elijo un i mayor o igual a cero tal que $xy^{i}z$ no pertenezca a L

De las cosas que me dan, no puedo asumir más de lo que se me dice.
De las cosas que elijo, puedo tomar las decisiones que crea convenientes.

Podemos tomar esto como una serie de pasos para probar que un lenguaje no cumple "*pumping*", con el objetivo de demostrar que dicho lenguaje no es regular. Piensenlo como una receta o un juego frente a un contrincante, en el cual este realiza los pasos 1 y 3, y ustedes los pasos 2 y 4.

Ejercicio

$L_1 = \{a^n b^n \mid n \geq 1\}$

1) Me dan un $p > 0$
2) Elijo una cadena α perteneciente a L con longitud mayor o igual a p

Elijo $\alpha = a^p b^p$, vale que $\alpha \in L_1$ y $|\alpha| \geq p$

3) Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$

Sin importar la descomposición dada, podemos asegurar que:

$$x = a^r \ (r \geq 0) \qquad y = a^t \ (t \geq 1) \qquad z = a^{p-r-t} b^p$$

pues $|xy| \leq p$ y los primeros p símbolos de α son todas a's.

4) Elijo un i mayor o igual a cero tal que $xy^{\text{i}}z$ no pertenezca a L

Elijo $i = 0$, de esta manera $xy^0z = xz = a^r.a^{p-r-t}b^p = a^{p-t}b^p$.
Como $t \geq 1$ puedo asegurar que $p - t \neq p$ y por lo tanto $xy^0z \notin L_1$

Ejercicio

$L_2 = \{a^n b^m \mid n \leq m\}$

Repito los 3 primeros pasos del ejercicio anterior:

1) Me dan un $p \geq 0$
2) Elijo $\alpha = a^p b^p$, vale que $\alpha \in L_2$ y $|\alpha| \geq p$
3) Podemos asegurar que $x = a^r$ $(r \geq 0)$, $y = a^t$ $(t \geq 1)$, $z = a^{p-r-t} b^p$

4) Elijo un i mayor o igual a cero tal que $xy^{\mathrm{i}}z$ no pertenezca a L

En este caso no puedo elegir $i = 0$ porque $xy^0z = xz \in L_2$.
Pero si tomo $i = 2$ tengo $\beta = xy^2z = xyyz = a^r a^t a^t a^{p-r-t} b^p = a^{p+t} b^p$.
Como $t \geq 1$ vale que $|\beta|_a = p + t > p = |\beta|_b$ y por lo tanto $xy^2z \notin L_2$.

Podría haber elegido también $i = 3$ (o cualquier $i \geq 2$), pero con encontrar uno sólo alcanza.
$i = 1$ nunca es una opción porque $xy^1z = xyz = \alpha \in L$.

Ejercicio

$L_3 = \{(ab)^n(cd)^n \mid n \geq 1\}$

Hay veces que es necesario (o conveniente) partir en casos. Hay que tener cuidado en cubrir **todos** los casos y para cada uno hacer la demostración.

1) Me dan un $p \geq 0$
2) Elijo $\alpha = (ab)^p(cd)^p$, vale que $\alpha \in L_3$ y $|\alpha| \geq p$
3) Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$

Existen 4 casos para y:

1. Empieza con a, termina con a
2. Empieza con a, termina con b
3. Empieza con b, termina con a
4. Empieza con b, termina con b

Ejercicio

$L_3 = \{(ab)^n(cd)^n \mid n \geq 1\}$

1. Empieza con a, termina con a
   Tenemos $x = (ab)^*,\ y = (ab)^*a,\ z = b(ab)^*(cd)^p$
   Tomamos $i = 2$, en $xyyz$ vamos a tener 2 a's seguidas.

2. Empieza con a, termina con b
   Tenemos $x = (ab)^*,\ y = (ab)^+,\ z = (ab)^*(cd)^p$
   Tomamos $i = 0$, en xz vamos a tener menos (ab) que (cd).

3. Empieza con b, termina con a
   Tenemos $x = (ab)^*a,\ y = (ba)^+,\ z = b(ab)^*(cd)^p$
   Tomamos $i = 0$, en xz vamos a tener menos (ab) que (cd).

4. Empieza con b, termina con b
   Tenemos $x = (ab)^*a,\ y = b(ab)^+,\ z = (ab)^*(cd)^p$
   Tomamos $i = 2$, en $xyyz$ vamos a tener 2 b's seguidas.

Ejercicio

$L_3 = \{(ab)^n(cd)^n \mid n \geq 1\}$

En este caso también podriamos haber hecho lo siguiente:

1) Me dan un $p \geq 0$
2) Elijo $\alpha = (ab)^p(cd)^p$, vale que $\alpha \in L_3$ y $|\alpha| \geq p$
3) Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$

Es condición necesaria para que $\beta \in L_3$ que $|\beta|_a = |\beta|_b = |\beta|_c = |\beta|_d = n$. Sabemos que y solo puede contener a's y b's (pero nunca c's o d's).

Si tomamos $i = 0$ entonces xz tendra al menos una a o una b menos, y por lo tanto la condición necesaria no se cumplirá. Algo similar sucede si tomamos $i = 2$, en este caso tendremos al menos una a o una b más.

Ejercicio

$L_4 = \{0^m 1^n \mid m > n \vee m \textit{ es par}\}$

Este es el ejercicio de la práctica. El inciso *a* pide mostrar que L_4 cumple el consecuente de pumping. Sin embargo, L_4 NO es regular (recuerden que no vale la vuelta de la implicancia). Por lo tanto, vamos a tener que recurrir a alguna(s) propiedad(es) para hacer la demostración.

Ejercicio

$L_{4a} = \{0^m 1^n \mid m > n \wedge m \text{ } es \text{ } impar\}$

Probemos primero el inciso a, tenemos que mostrar que existe un p tal que para toda cadena del lenguaje que cumple $|\alpha| \geq p$ existe una descomposición xyz con $|xy| \leq p$ y $|y| \geq 1$ tal que que $\forall i \in N, xy^iz \in L_4$

Tomemos $p = 2$, veamos los casos que cumplen las cadena $\alpha \in L_4$ de largo 2 o más:

1. $\alpha = 1^k$, dado que $\alpha \in L_4$ entonces $|\alpha_4|$ es par. Tomamos $y = 11$ y es trivial ver que $xy^iz \in L_4$ para todo i
2. $\alpha = 0^{2k}1^l$, procedemos de manera similar, tomando $y = 00$, al bombear la cantidad de 0s seguirá siendo par
3. $\alpha = 0^{2k+1}1^l$, con $k > 0$ dado que $\alpha \in L_4$ y $|\alpha| \geq 2|$, entonces sabemos que $2k + 1 > l$, tomamos $x = \lambda, y = 0$. Ahora tenemos que ver que $\forall i$ vale que $xy^iz \in L_4$, tenemos 2 casos:
    - $i = 0$, entonces $|xy^0z|_0 = |\alpha|_0 - 1$, es decir que es par y por lo tanto $xy^0z \in L_4$
    - $i \geq 1$, para todo i vamos a estar agregando 0s pero no 1s, por lo tanto va a seguir valiendo la propiedad de $m > n$ y la cadena va a estar en L_4

Ejercicio

$L_{4a} = \{0^m 1^n \mid m > n \land m \text{ es impar}\}$

Probemos ahora que L_4 es no-regular.
Importante: es cierto que $\{0^m 1^n \mid m > n\}$ es no-regular y $\{0^m 1^n \mid m \text{ es par}\}$ es regular, pero la unión **NO** es disjunta (contraejemplo: $\alpha = 001$), por lo tanto no se puede usar esa propiedad.

Planteamos $L_{4a} = \{0^m 1^n \mid m > n \land m \text{ es impar}\}$, $L_{4b} = \{0^m 1^n \mid m \text{ es par}\}$:

- Vale que $L_4 = L_{4a} \overset{d}{\cup} L_{4b}$
- Es fácil ver que L_{4b} es regular (con la ER $(00)^* 1^*$)
- Si probamos que L_{4a} no es regular, podemos usar la propiedad de union disjunta para demostrar que L_4 tampoco es regular

Ejercicio

$L_{4a} = \{0^m 1^n \mid m > n \wedge m \text{ es impar}\}$

Tener un AND ($\wedge$) es más fácil que un OR ($\vee$) ya que podemos elegir α acordemente y luego ver que xy^iz *rompa* **al menos una** de las condiciones (en un OR deberíamos *romper* todas las condiciones).

1. Me dan un $p \geq 0$
2. Elijo $\alpha = 0^{2p+1}1^{2p}$, vale que $\alpha \in L_{4a}$ y $|\alpha| \geq p$
3. Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$
   Sabemos que $x = 0^r,\ y = 0^t\ (t \geq 1),\ z = 0^{2p+1-r-t}1^{2p}$
4. No vamos a poder *romper* la imparidad, pero sí la otra condición.
   Tomando $i = 0$ tenemos $xy^0z = xz = 0^r0^{2p+1-r-t}1^{2p} = 0^{2p+1-t}1^{2p}$
   y luego $t \geq 1 \Rightarrow 2p + 1 - t \leq 2p \Rightarrow xy^0z \notin L_{4a}$

Por lo tanto L_{4a} no es regular y entonces L_4 tampoco lo es $\square$

Ejercicio

$L_4 = \{0^m 1^n \mid m > n \vee m \textit{ es par}\}$

Otra alternativa es utilizar la propiedad del complemento.

Importante: ${L_4}^c \neq \{0^m 1^n \mid m \leq n \wedge m \textit{ es impar}\}$ (pensar en $\beta = 10$)

No hace falta (y suele generar problemas) escribir ${L_4}^c$. Si buscan a L en

1. Me dan un $p > 0$
2. Elijo una cadena α perteneciente a L con longitud mayor o igual a p
3. Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$
4. Elijo un i mayor o igual a cero tal que $xy^{\mathrm{i}}z$ no pertenezca a L

verán que todas las menciones son sobre la pertenencia,
y sabemos que $\alpha \in L \iff \alpha \notin L^c$.

Ejercicio

${L_4}^c$ siendo $L_4 = \{0^m 1^n \mid m > n \vee m \text{ es par}\}$

1. Me dan un $p \geq 0$
2. Elijo $\alpha = 0^{2p+1} 1^{2p+1}$, vale que $\alpha \in {L_4}^c$ (pues $\alpha \notin L_4$) y $|\alpha| \geq p$
3. Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$
   Sabemos que $x = 0^r,\ y = 0^t\ (t \geq 1),\ z = 0^{2p+1-r-t} 1^{2p+1}$
4. Tomando $i = 2$ tenemos $xy^2z = xyyz = 0^{2p+1+t} 1^{2p+1}$
   $t \geq 1 \Rightarrow 2p + 1 + t > 2p + 1 \Rightarrow xy^2z \in L_4 \Rightarrow xy^2z \notin {L_4}^c$

Por lo tanto ${L_4}^c$ no es regular y entonces L_4 tampoco lo es $\square$

Ejercicio

$L_5 = \{a^n b^m \mid n \neq m\}$

Partir de una cadena $\alpha \in L_5$ y luego encontrar un i tal que $xy^iz \notin L_5$ parece algo complicado (es posible, pero involucra utilizar cosas *raras* como la función factorial $p!$).

Es mucho más fácil partir de algo igual y luego llegar a algo distinto! Para esto podemos usar nuevamente la propiedad del complemento.

Recuerden que intentar escribir ${L_5}^c$ puede ser más dañino que beneficioso. Sin embargo, en este caso pueden usar exactamente la misma demostración que hicimos para $L_1 = \{a^n b^n \mid n \geq 1\}$ para mostrar que ${L_5}^c$ no es regular. Luego, por la propiedad, L_5 tampoco es regular.

Ejercicio

$L_6 = \{a^c \mid c \text{ es cuadrado perfecto}\}$

Un número cuadrado perfecto es un número entero que es el cuadrado de algún otro. Algunas cadenas que pertenecen a L_6 son a^4, a^9, a^{16}.

1. Me dan un $p \geq 0$
2. Elijo $\alpha = a^{p^2}$, vale que $\alpha \in L_5$ y $|\alpha| \geq p$
3. Me dan una descomposición $\alpha = xyz$ con $|xy| \leq p$ y $|y| \geq 1$
4. Con $i = 2$ vale lo siguiente:
   - $|xy^2z| = |xyz| + |y| \leq p^2 + p < p^2 + 2p + 1 = (p+1)^2$
   - $|xy^2z| = |xyz| + |y| > |xyz| = p^2$ pues $|y| \geq 1$
   - Por lo tanto $p^2 < |xy^2z| < (p+1)^2$
   - $|xy^2z|$ no puede ser entonces un cuadrado perfecto, es decir $xy^2z \notin L_6$

Para tener en cuenta

- Tratar de entender cual es el lenguaje, para elegir la cadena adecuadamente.
- Si con la cadena elegida no se encuentra ningún i, volver hacia atras y elegir otra α.
- Pueden haber más de una cadena y mas de un i que sirvan como elección para hacer verdadera la fórmula. Alcanza con encontrar uno.
- No podemos asumir nada sobre la descomposición más allá de $|xy| \leq p$ y $|y| \geq 1$.
- Recordar que no vale la vuelta: que cumpla "*pumping*" no implica que el lenguaje sea regular.
- Podemos utilizar las propiedades de los lenguajes regulares vistas para facilitar nuestra prueba.

¿Preguntas?