Lema del bombeo

Autómatas y Lenguajes Formales

24 de septiembre de 2022

Lema del bombeo

Lema 1 *Sea* $\mathbf{A}$ *un conjunto regular (un conjunto de cadenas regulares), entonces se tiene las propiedades para* $\mathbf{A}$*: (P) Existe* $k \geq 0$ *tal que para cualesquiera cadenas* x, y, z *con* $xyz \in \mathbf{A}$ *y* $|y| \geq k$ *existen cadenas* u, v, w *que cumplen* $y = uvw$*,* $v \neq \epsilon$ *y para toda* $i \geq 0$ *la cadena* $xuv^iwz \in \mathbf{A}$.

Este lema lo que te asegura es que si no cumple esa propiedad el lenguaje no es regular, pero **ojo**, que se cumpla no implica que el lenguaje sea regular. Si un lenguaje es regular la mejor manera de demostrarlo es construyendo su autómata finito. Realmente lo que usamos para demostrar que un lenguaje no es regular es la negación de este lema.

Para demostrar que un lenguaje no es regular hacemos algo como un juego de apuesta con una entidad desconocida (en el libro de Kozen dice que el demonio, pero realmente lo que trata de dar a entender es que tu no controlas unas partes del proceso). El juego es el que sigue:

El demonio apuesta contigo, él dice que un lenguaje $\mathbf{A}$ dado es regular, tú debes demostrar que no lo es, empezamos por turnos y el demonio como es mañoso empieza primero:

1. El demonio elige la k (como le está apostando a que el lenguaje es regular, lo más sensato es que elija k igual al número de estados del autómata finito determinista).
2. Tú eliges x, y, z a conveniencia, tales que $xyz \in \mathbf{A}$ y $|y| \geq k$.
3. Ahora le toca al demonio y escoge u, v, w con $y = uvw$ y $v \neq \epsilon$, es decir, v no puede ser la cadena vacía.
4. Tu jugada de remate, para ganarle al demonio, es elegir la $i \geq 0$ que demostrará que el lenguaje no es regular.

Aquí ganas si $xuv^iwz \notin \mathbf{A}$.

Veámoslo con un ejemplo: *Demuestra que el lenguaje* $\mathbf{A} = \{a^nb^n | n \geq 0\}$ *no es regular*.

Vamos con los pasos:

1. Te dan una $k \geq 0$ cualquiera.

2. Tú eliges $x = a^k$, $y = b^k$ y $z = \epsilon$, lo cual se vale ya que sobre x y z no hay restricción, una o ambas pueden ser la cadena vacía, y se cumple que $|a^k| = k \geq k$, el tamaño de la cadena de a's k-veces es k, la cual cumple la parte de la igualdad. Ademas $xyz = a^k b^k \in \mathbf{A}$.

3. Tu contrincante ya vio tus negras intenciones y no se va a dejar, elige unas u, v, w tales que $|u| = j$, $|v| = m$ y $|w| = n$ (se vale porque entonces $v \neq \epsilon$ siempre que $m > 0$). Para que se cumpla con lo que tu propusiste entonces $j+m+n = |y| = k$, es decir, la longitud sumada de las tres cadenas debe ser igual a la longitud de la cadena y.

4. Aquí lo rematas, sin importar como sean las cadenas que eligió el demonio en el paso 3, sin importar los valores de sus longitudes, si eliges $i = 2$ le ganas, ya que con ese valor la cadena total sería $xuv^iwz = a^k b^j (b^m)^2 b^n = a^k b^{j+m+m+n} = a^k b^{k+m} \notin \mathbf{A}$.

Con la elección del demonio llegas a que si se cumpliera el lema del bombeo habría una cadena que no está en el lenguaje, es decir, no es regular.

Un ejemplo más: *Muestra que $L = \{a^p | p$ es primo$\}$ no es regular.*

Lo hacemos por pasos como hemos visto que es el juego propuesto por Kozen.

- El demonio elige k, podemos sospechar que es el número de estados del autómata.

- Nos toca, sea $p > k$ un número primo. Elegimos $y = a^p$, como pueden ver $y \in L$.

- El demonio propone u, v, w tales que $|uv| \leq k$ y $|v| > 0$ como dictan las reglas, pero no nos da la descripción exacta. Podemos suponer que $v = a^m$ con $1 \leq m \leq k$, sin perder ninguna generalidad.

- Ahora nos toca dar el movimiento final para ganar la discuisión, proponemos $i = p + 1$, entonces podemos ver que

$$
\begin{aligned}
|uv^iw| &= |uvw| + |v^{i-1}| \\
&= p + (i-1)|v| \\
&= p + (i-1)m \text{ (ya que supusimos } v = a^m) \\
&= p + pm \text{ (como propusimos } i = p+1) \\
&= p(1+m) \text{ (si } m \geq 1, 1+m \geq 2)
\end{aligned}
$$

Ese último valor corresponde a la longitud de $|y| = |uvw| = p(1+m)$ que de cumplir ser parte del lenguaje debería ser un número primo, pero esa longitud no es un número primo, por lo tanto $|uv^iw| \notin L$, entonces L no es un lenguaje regular■.

Más ejemplos:

Muestra que $L = \{a^n b^n a b^{n+1}\}$ no es regular.

Como siempre, vamos por pasos:

- El demonio elige k, como es costumbre supones que es la cantidad de estados del autómata finito asociado.

- Proponemos $y = a^k b^k a b^{k+1}$, de tal forma que $|y| = (3k+2) > k$

- El demonio propone u, v, w tal que $y = uvw$, con $|v| > 0$.
- Debemos proponer una i yal que $uv^iw \notin L$. Hay opciones:
  - $v = a^r$
  - $v = b^s$
  - $v = b^r a$
  - ab^r
- Para el caso $y = a^r$, tomamos $i = 0$

$$y = uvw = (a^{k-r})(a^r)(b^k ab^{k+1}) \quad (1)$$

$$uv^0w = a^{k-r}b^k ab^{k+1} \quad (2)$$

- Para el caso $v = b^s$ tomamos $i = 0$

$$\begin{aligned} y =& uvw = (a^k b^{k-s})(b^s)ab^{k+1} \\ uv^iw =& a^k b^{k-1}ab^{k+1} \text{ ó} \\ y =& uvw = a^k b^k a(b^s)b^{k-s+1} \\ uv^iw =& a^k b^k ab^{k-s+1} \end{aligned}$$

- Para el caso $v = b^r a$ tomamos $i = 2$

$$\begin{aligned} y =& uvw = (a^k b^{k-r})(b^r a)b^{k+1} \\ uv^2w =& a^k b^k ab^k ab^{k+1} \end{aligned}$$

- Para el caso $v = ab^r$ tomamos $i = 2$

$$\begin{aligned} y =& uvw = a^k b^k(ab^r)b^{k-r+1} \\ uv^2w =& a^k b^k ab^r ab^{k+1} \end{aligned}$$

No cumplen en todos los casos estar en el lenguaje, no es regular ■.
Otro ejemplo, que dejare más corto, espero ya quedara claro el anterior:
Demuestra que el lenguaje $\mathbf{A} = \{a^{2^n} | n \geq 0\}$ *(las cadenas donde* a *aparece repetida en potencias de dos, como* ϵ*, nn, nnnn y nnnnnnnn) no es regular.*

1. $k \geq 0$
2. $x = z = \epsilon$, $y = a^{2^k}$, ya que $xyz = a^{2^k} \in \mathbf{A}$ y $|y| = 2^k \geq k$.
3. u, v, w tales que $|u| = j$, $|v| = m$ y $|w| = n$, que cumplen $|m| > 0$ y $j+m+n = 2^k$ y al menos $|u| \neq \epsilon$ o $|u| \neq \epsilon$.
4. Escoges cualquier $i > 1$ y le ganas, ya que queda la cadena total $|xuv^iwz| = j + im + n = j + m + n + (i-1)m = 2^k + (i-1)m$ y las cadenas de esa longitud no están en el lenguaje ¿porqué?

   Sabemos que $m < 2^k$ ya que $j + m + n = 2^k$, a su vez $(i-1)m < (i-1)2^k$, por lo tanto $2^k < 2^k + (i-1)m < 2^k + (i-1)2^k = i2^k$. Podemos tomar el primer caso, $i = 2$, entonces $2^k < 2^k + (i-1)m < 2 \times 2^k = 2^{k+1}$. Entonces la longitud de la cadena ya bombeada está entre 2^k y 2^{k+1}, pero no puede ser que esa magnitud ser de la forma 2^s ya que no hay un natural s entre k y $k+1$ ■

Hace falta que vean el teorema de Myhill-Nerode, pero ese lo dejo para que ustedes lo chequen, se puede también probar que un lenguaje no es regular a partir de él. Revísenlo y será bueno chequen estos ejemplos en el marco de este teorema.

Referencias

[1] Kozen, Dexter C. "Automata and Computability" Springer (1997)

[2] Sipser, Michael "Introduction to the Theory of Computation" 2a ed., Thomson Course Tecnology (2006)

[3] Kandar, Shyamalendu "Introduction to Automata Theory, Formal Languages and Computation", Dorling Kindersley (India) Pvt. Ltd. (2013)