Autómatas y Lenguajes Formales
Nota 12. Lema del bombeo y propiedades de cerradura para lenguajes libres del contexto*

Noé Salomón Hernández S.

Hay un lema del bombeo para LLC similar al que se tiene para lenguajes regulares. Puede ser usado del mismo modo para demostrar que ciertos lenguajes no son libres del contexto.

1. Lema del bombeo para LLC

Teorema 1.1 *Para cada LLC A, existe $k \geq 0$ tal que toda cadena $z \in A$ de tamaño al menos k puede descomponerse en cinco subcadenas $z = uvwxy$ tal que $vx \neq \varepsilon$, $|vwx| \leq k$, y para toda $i \geq 0$, $uv^iwx^iy \in A$.*

Demostración. La demostración se dará en clase. $\dashv$

El lema del bombeo anterior es muy útil en su forma contrapositiva. Dicha forma nos permite conlcuir que A **no** es libre del contexto al cumplirse la siguiente propiedad.

Propiedad 1.2 Para toda $k \geq 0$, existe $z \in A$ de tamaño al menos k tal que para toda forma de descomponer z en las subcadenas $z = uvwxy$ con $vx \neq \varepsilon$ y $|vwx| \leq k$, existe una $i \geq 0$ tal que $uv^iwx^iy \notin A$.

La propiedad anterior es equivalente a decir que se tiene una estrategia ganadora para la siguiente interacción que simula una partida contra un adversario.

1. El adversario elige $k \geq 0$.
2. Exhibimos una $z \in A$ de tamaño al menos k.
3. El adversario elige las cadenas u, v, w, x y y tales que $z = uvwxy$, $|vx| > 0$, y $|vwx| \leq k$.
4. Tomamos una $i \geq 0$. Si $uv^iwx^iy \notin A$, entonces ganamos.

Si deseamos demostrar que un lenguaje A no es libre de contexto, basta con exhibir una estrategia ganadora para la interacción anterior, es decir, no importa lo que el adversario realice en los pasos 1 y 3, tenemos la jugada correcta en los pasos 2 y 4 para derrotarlo.

*Esta nota se basa en las notas del prof. Rajeev y en los libros *Introduction to Automata Theory, Languages, and Computation*; *Automata and Computability*; *Introduction to Languages and the Theory of Computation* y *Introduction to the Theory of Computation*.

Ejemplo 1.3 Usemos el lema del bombeo para demostrar que el conjunto

$$A = \{a^n b^n c^n \mid n \geq 0\}$$

no es libre del contexto. Para esto veremos que siempre podemos ganar la interacción contra el adversario del siguiente modo:

1. Digamos que el adversario elige k en el paso 1. Nuestra estrategia ganadora debe ser válida para cualquier k.

2. Consideramos $z = a^k b^k c^k$. Sabemos que $z \in A$ y $|z| = 3k \geq k$.

3. Ahora digamos que el adversario elige u, v, w, x, y tal que $z = uvwxy$, $vx \neq \varepsilon$, y $|vwx| \leq k$.

4. Tomamos $i = 2$. Consideramos dos casos, dependiendo de si las subcadenas v y x contienen más de un tipo de símbolo del alfabeto.

   a) v y x contienen ambas un único tipo de símbolo del alfabeto, ya sea a, b ó c, no necesariamente el mismo, es decir, es posible que tanto v como x contengan el mismo tipo de símbolo, o que una subcadena contenga un tipo de símbolo y la otra subcadena un tipo diferente, considerando a lo más dos símbolos distintos. De esta manera, v y x no cubren los tres tipos diferentes de símbolos. La cadena uv^2wx^2y no puede tener un número igual de a's, de b's y de c's. Por lo tanto no es un miembro de A.

   b) Si v o x contiene más de un tipo de símbolo, entonces uv^2wx^2y puede tener un número igual de los tres tipos de símbolos pero mostrando un patrón equivocado, distinto a $a^*b^*c^*$. Por lo tanto no es un miembro de A.

   En ambos casos, podemos asegurar que $uv^2wx^2y \notin A$. Hemos exhibido una estrategia ganadora. Por el lema del bombeo, A no es libre del contexto.

Ejemplo 1.4 Usemos el lema del bombeo para demostrar que el conjunto

$$A = \{a^n b^m a^n b^m \mid m, n \geq 0\}$$

no es libre del contexto. Para esto veremos que siempre podemos ganar la interacción contra el adversario del siguiente modo:

1. Digamos que el adversario elige k en el paso 1.

2. Consideramos $z = a^k b^k a^k b^k$. Llamamos a cada subcadena de la forma a^k o b^k un bloque. Sabemos que $z \in A$ y $|z| = 4k \geq k$.

3. Ahora digamos que el adversario elige u, v, w, x, y tal que $z = uvwxy$, $vx \neq \varepsilon$, y $|vwx| \leq k$.

4. Tomamos $i = 2$.

   a) Si el adversario optó por que v o x contengan tanto a's como b's, es decir, si v o x toman elementos de ambos lados de la frontera de un bloque, entonces uv^2wx^2y no es de la forma $a^*b^*a^*b^*$, y no forma parte de A.

   b) Si v y x se encuentran ambos en el mismo bloque, entonces uv^2wx^2y tiene un bloque más grande que los otros tres, por lo tanto no forma parte de A.

c) Si v y x pertenecen a distintos bloques, los bloques deben ser contiguos; esto es por la restricción $|vwx| \leq k$ que puede pensarse como una ventana corrediza sobre la cadena $z = a^k b^k a^k b^k$, por el tamaño k de la ventana no hay manera que ésta se extienda a través de tres bloques contiguos de z, únicamente alcanza dos. De modo que uno de los bloques abarcando v o x debe ser un bloque de a's y el otro un bloque de b's. Entonces uv^2wx^2y tiene dos bloques de a's de diferente tamaño y dos bloques de b's de diferente tamaño.

En todos los casos, podemos asegurar que $uv^2wx^2y \notin A$, así hemos exhibido una estrategia ganadora. Por el lema del bombeo, A no es libre del contexto.

Algunas veces, distintos casos del paso 4 requieren distintas elecciones de i para lograr que la cadena bombeada no forme parte del lenguaje. Esto se ejemplifica enseguida.

Ejemplo 1.5 Sea

$$A = \{s \in \{a, b, c\}^* \mid n_a(s) < n_b(s) \text{ y } n_a(s) < n_c(s)\}$$

recuerde que $n_\sigma(s)$ denota el número de presencias del símbolo σ en la cadena s. Vamos a demostrar, por medio de la interacción con el adversario, que A no es libre del contexto. Notemos primero que los elementos de A son las cadenas en las que hay más b's que a's y hay más c's que a's.

1. El adversario elige k en el paso 1.
2. Consideramos $z = a^k b^{k+1} c^{k+1}$. Sabemos que $z \in A$ y $|z| \geq k$.
3. Ahora digamos que el adversario elige u, v, w, x, y tal que $z = uvwxy$, $vx \neq \varepsilon$, y $|vwx| \leq k$.
4. Analicemos los siguientes casos.

   a) Si vx contiene al menos una a, entonces vx no contiene una c puesto que la ventana que define $|vwx|$ es de tamaño a lo más k sobre $z = a^k b^{k+1} c^{k+1}$, dicha ventana tiene un extremo anclado en el bloque a^k y el otro extremo no llega al bloque c^{k+1}. Con $i = 2$, resulta que uv^2wx^2y tiene al menos $k+1$ a's y exactamente $k+1$ c's. Así, $uv^2wx^2y \notin A$.

   b) Si vx no contiene una a, entonces contiene al menos una b ó al menos una c. En este caso, tomamos $i = 0$, de modo que uv^0wx^0y contiene exactamente k a's, y no más de k b's ó no más de k c's. Así $uv^0wx^0y \notin A$.

Podemos encontrar valores del bombeo, i, tal que dada cualquier descomposición del adversario logramos que $uv^iwx^iy \notin A$. Hemos exhibido una estrategia ganadora. Por el lema del bombeo, A no es libre del contexto.

2. Propiedades de cerradura para los LLC

Teorema 2.1 *Los lenguajes libres del contexto (LLC) son cerrados bajo la unión, concatenación y estrella de Kleene.*

Demostración. Sean L_1 y L_2 dos lenguajes libres del contexto, con gramáticas G_1 y G_2, respectivamente. Supongamos que todas las variables de G_1 y de G_2 son distintas, con símbolo terminal S_1 y S_2, respectivamente.

Formamos las nuevas gramáticas para

a) $L_1 \cup L_2$ al combinar G_1 y G_2 a través de la producción $S \longrightarrow S_1 \,|\, S_2$.

b) $L_1 \bullet L_2$ al combinar G_1 y G_2 a través de la producción $S \longrightarrow S_1S_2$.

c) L_1^* agregando la producción $S \longrightarrow S_1S \,|\, \varepsilon$.

$\dashv$

Teorema 2.2 *Los LLC* **no** *son cerrados bajo la intersección.*

Demostración. Cosideremos los lenguajes libres del contexto

$$L_1 = \{a^ib^jc^k \,|\, i = j, \;\; i,j,k \geq 0\} \quad \text{y} \quad L_2 = \{a^ib^jc^k \,|\, j = k, \;\; i,j,k \geq 0\}.$$

Definimos ahora,

$$\begin{aligned} L = L_1 \cap L_2 &= \{a^ib^jc^k \,|\, i = j \text{ y } j = k, \;\; i,j,k \geq 0\} \\ &= \{a^nb^nc^n \,|\, n \geq 0\} \end{aligned}$$

sin embargo, demostramos arriba mediante el lema del bombeo que L no es libre del contexto. $\dashv$

Teorema 2.3 *Los LLC* **no** *son cerrados bajo complemento.*

Demostración. Supongamos que los LLC sí son cerrados bajo complemento. Sean L_1 y L_2 dos LLC. Tenemos entonces que el siguiente lenguaje es LLC

$$\overline{\overline{L_1} \cup \overline{L_2}} = L_1 \cap L_2$$

pero aquí estaríamos diciendo que la intersección es una operación cerrada para los LLC, lo cual no es cierto. Suponer que los LLC son cerrados bajo complemento nos llevó a una contradicción. Por lo tanto, los LLC **no** son cerrados bajo complemento. $\dashv$

Teorema 2.4 *Los LLC son cerrados bajo la operación de reversa.*

Demostración. Hay que revertir el lado derecho de cada producción. Por ejemplo, $A \longrightarrow BCD$ quedaría como $A \longrightarrow DCB$. $\dashv$

Teorema 2.5 *Los LLC son cerrados bajo la intersección con lenguajes regulares, es decir, si L es un LLC y R es un lenguaje regular, entonces $L \cap R$ es LLC.*

Demostración. La idea es ejecutar un autómata finito en paralelo con un PDA, cuyo resultado sea otro PDA, como se sugiere en la Figura 1.

Sea $P = (Q_P, \Sigma, \Gamma_P, \delta_P, p_0, Z_0, F_P)$ un PDA que reconoce L por estados finales y sea $D = (Q_D, \Sigma, \delta_D, q_0, F_D)$ un AFD para R, construimos un autómata de pila M para $L \cap R$ el cual toma como estados las parejas ordenadas (q, p), con $q \in Q_D$ y $p \in Q_P$. Formalmente, tenemos $M = (Q_D \times Q_P, \Sigma, \Gamma_P, \delta_M, (q_0, p_0), Z_0, F_M)$ donde

$$F_M = \{(q, p) \in F_D \times F_P\},$$

y $\delta_M((q, p), a, X)$ se define como el conjunto de pares $((r, s), \gamma)$ tal que:

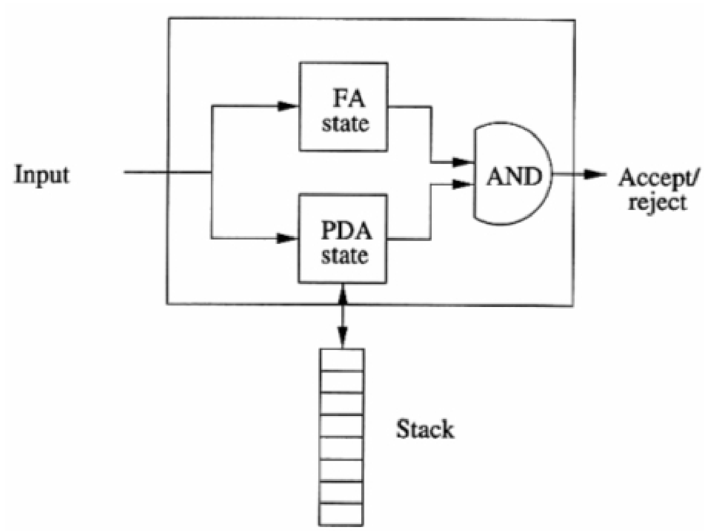

Figura 1: Un PDA y un autómata finito ejecutándose en paralelo para crear un nuevo PDA.

1. $r = \widehat{\delta_D}(q, a)$,
2. $(s, \gamma) \in \delta_P(p, a, X)$.

Por cada movimiento del PDA P, realizamos el mismo movimiento en el PDA M, además de que llevamos acarreando el estado del AFD D en la primer componente del estado de M. Observemos que si $a = \varepsilon$, entonces $r = \widehat{\delta_D}(q, a) = q$, es decir, no hay cambio de estado en D cuando P realiza una transición ε.

Una inducción sencilla sobre el número de movimientos hechos por los PDAs nos permite concluir que $(p_0, w, Z_0) \vdash_P^* (p, \varepsilon, \gamma)$ si y solo si $((q_0, p_0), w, Z_0) \vdash_M^* ((q, p), \varepsilon, \gamma)$, donde $q = \widehat{\delta_D}(q_0, w)$. Como (q, p) es un estado final de M si y solo si p es un estado final de P y q es un estado final de D, concluimos que M acepta a w si y solo si tanto P como D la aceptan, i.e., $w \in L \cap R$. $\dashv$

Ejemplo 2.6 ¿Es el lenguaje $A = \{ww \mid w \in \{a, b\}^*\}$ libre de contexto? Supongamos por un momento que sí y consideremos el lenguaje siguiente:

$$\begin{aligned} A' &= A \cap L(a^*b^*a^*b^*) \\ &= \{a^n b^m a^n b^m \mid m, n \geq 0\}. \end{aligned}$$

Por el Teorema 2.5, A' es un lenguaje libre de contexto, pero esto no es posible porque en el Ejercicio 1.4 demostramos que A' no es libre de contexto. Luego, suponer que A es libre de contexto nos condujo a esta situación contradictoria; por consiguiente, A no es libre de contexto.