El lema de bombeo para lenguajes regulares

Lenguajes, Gramáticas y Autómatas, **cuarto cuatrimestre (primavera) de Ingeniería en Informática**
http://webdiis.unizar.es/asignaturas/LGA

Rubén Béjar Hernández, Pedro Javier Álvarez
Dpto. Informática e Ingeniería de Sistemas
Universidad de Zaragoza

Introducción

El lema de bombeo para lenguajes regulares enuncia una propiedad que cumplen todos los lenguajes regulares infinitos (y también algunos lenguajes que no son regulares). Gracias a este lema podremos demostrar que ciertos lenguaje infinitos no son regulares. Es importante hacer notar que el lema de bombeo es una herramienta adecuada para demostrar que un lenguaje **no** es regular, pero no lo será para demostrar que un lenguaje **si** es regular (por el hecho de que existen algunos lenguajes no regulares que la cumplen). Por tanto, si un lenguaje no cumple el lema de bombeo no es regular, pero si lo cumple no podremos decir si es o no regular.

Enunciado del Lema de Bombeo

Para todo lenguaje regular infinito L, existe una constante n, dependiente de ese lenguaje, de forma que si w es una cadena de L con $|w| \geq n$, podemos partir w en tres cadenas, x, y, z, de forma que:

- $w = xyz$,
- $y \neq \varepsilon$ (o dicho de otro modo, que $|y| \geq 1$),
- $|xy| <= n$
- Para cualquier $k \geq 0$, la cadena xy^kz pertenece a L.

Más formalmente:

$\forall$ lenguaje regular infinito L sobre un alfabeto $\sum$
$\exists\, n \in N$ /
$\forall\, w \in L$ / $|w| \geq n$
$\exists\, x, y, z \in \sum^*$ / $w = xyz$, $y \neq \varepsilon$, $|xy| <= n$,
$\forall\, k \geq 0$, $xy^kz \in L$

O sea que para cualquier cadena de L lo bastante larga, siempre podremos encontrar una partición en tres subcadenas, con una no vacía en el medio (la y) que no está demasiado lejos del comienzo de la palabra, que podremos "bombear"; es decir, que si se repite la subcadena y cualquier número de veces, la cadena resultante también pertenecerá a L.

Demostración de que un lenguaje no es regular

Dado que para todo lenguaje regular infinito se cumple el lema de bombeo, si nos dan un lenguaje infinito y demostramos que para él no se cumple, habremos demostrado que no es un lenguaje regular. Como el lema de bombeo es una propiedad que se cumple para todas las cadenas de longitud mayor o igual a cierta n, bastará encontrar una cadena de ese lenguaje, de longitud mayor o igual a esa n, que no se pueda "bombear" para demostrar que el lenguaje no es regular. Con esta idea en mente, los pasos a dar para demostrar que un lenguaje dado no es regular son los siguientes:

1. Elegir una palabra w que pertenezca al lenguaje dado. Podemos elegir cualquier palabra del lenguaje, pero debe ser una cuya longitud sea mayor o igual que una constante n que desconocemos (la constante del lema de bombeo). Como desconocemos n, lo habitual será elegir una palabra en función de un n cualquiera y cuya longitud sea mayor o igual que n.
2. El lema de bombeo dice que si el lenguaje fuera regular, podríamos encontrar una forma de partir esa palabra w en tres, cumpliendo ciertas restricciones, y que esa partición sería bombeable. Como queremos demostrar que el lenguaje no es

regular, tendremos que demostrar que no hay ninguna forma de partir la palabra en tres cumpliendo las restricciones del lema, y que después se pueda bombear siempre.

3. Finalmente bastará con encontrar una constante $k \geq 0$ que haga que ninguna de las particiones posibles de w sea bombeable.

Más formalmente, para demostrar que un lenguaje L sobre un alfabeto Σ no es regular habrá que demostrar que:

$\forall\ n \in N$
$\exists\ w \in L\ /\ |w| \geq n,$
$\forall\ x, y, z \in \Sigma^*\ /\ w = xyz,\ y \neq \varepsilon,\ |xy| <= n,$
$\exists\ k \geq 0\ /\ xy^kz \notin L$

Es importante hacer notar que en esta demostración, por reducción al absurdo, de que un lenguaje no es regular, los cuantificadores existenciales “para todo” y “existe” están alternados al revés que en el enunciado del lema. Esto es, intuitivamente, porque estamos demostrando lo contrario que el lema de bombeo: éste enuncia una propiedad que cumplen todos los lenguajes regulares y la demostración precedente demuestra que un lenguaje no es regular, o sea que no cumple esa propiedad.

Ejemplo de demostración de que un lenguaje no es regular

Sea el lenguaje $L = \{a^{2n}b^n \mid n \geq 0\}$. Demostrar que L no es regular.

$\boxed{1.}$

Voy a suponer que el lenguaje es regular. Si lo es, y como es infinito, para él se cumplirá el lema de bombeo. Sea por tanto $n \in N$ la constante del lema de bombeo para L (constante que no conozco).
Elijo una palabra que pertenezca a L y de longitud mayor o igual a n:
$w = a^{2n}b^n$, tenemos que $w \in L$ y $|w| = 3n$ y por tanto $|w| \geq n$, sea cual sea n.

$\boxed{2.}$

Tengo que encontrar todas las formas de partir la palabra elegida w en tres xyz que cumplan las restricciones del lema de bombeo:

- $w = xyz$
- $y \neq \varepsilon$
- $|xy| <= n$ (siendo n la constante del lema)

Si me fijo en la palabra w que he elegido, cualquier x,y,z que cumplan las condiciones (restricciones) del lema serán de la siguiente forma:

- $x = a^i$
- $y = a^j$
- (con $j \geq 1$ puesto que $y \neq \varepsilon$, con $i+j <= n$ puesto que $|xy| <= n$)
- $z = a^{2n-i-j}b^n$

Se que la x y la y estarán formadas sólo por aes porque la palabra w que he elegido, y que estoy partiendo, tiene 2n aes al principio y la longitud de xy es menor o igual que n. También se las restricciones que cumplen sus índices (i, j) porque me las impone el lema de bombeo. Se puede ver además que obviamente $w = xyz$, porque $xyz = a^i\, a^j\, a^{2n-i-j}b^n = a^{2n}b^n = w$.

$\boxed{3.}$

Debo encontrar ahora una constante $k \geq 0$ con la que ninguna de las posibles particiones de w que hemos encontrado en el punto anterior sea bombeable.
Si elijo k = 2 y bombeo las x, y, z encontradas en el punto anterior para esa constante, tendré que:

$xy^2z = a^i a^{2j} a^{2n-i-j}b^n = a^{2n+j}b^n$ (para cualquier i y j, o sea para cualquiera de las particiones "legales" de la w elegida según el lema de bombeo).

Pero como $j \geq 1$ (ver punto 2, es una de las restricciones del lema) tengo que $xy^2z = a^{2n+j}b^n$, es una palabra que no pertenece al lenguaje porque tiene más del doble de aes que de bs (al menos una a más). He llegado por tanto a una contradicción (una palabra que no es bombeable de ninguna forma para al menos una constante k en un lenguaje supuestamente regular) que viene de suponer precisamente que el lenguaje L es regular, luego L no es regular.

Errores habituales al intentar la demostración anterior

Esta sección contiene varias soluciones incorrectas, propuestas por varios alumnos como parte del examen de 2ª convocatoria de LGA del curso 2002-03, al ejemplo de demostración anterior. El objetivo que se persigue es poder explicar algunos de los errores que se cometen con frecuencia. Para lograrlo, se detalla explícitamente el contenido para cada paso de la demostración (texto en cursiva) de algunas soluciones propuestas y, posteriormente, se describen los distintos errores que contienen.

- **<u>Ejemplo número 1:</u>**

$\boxed{1.}$
Tomaré como constante del lema de bombeo p=2n.
Elijo la palabra $w = a^n b^{n/2}$, $|w| = n + n/2 > n$.

(Los pasos $\boxed{2.}$ y $\boxed{3.}$ de la demostración son irrelevantes.)

Errores:
1. No se especifican cuáles son las suposiciones iniciales sobre las que se va a fundamentar la demostración (¿el lenguaje es regular?, ¿qué implica el hecho que se suponga que sea regular?,...)
2. El lenguaje especifica que $n \geq 0$. Esto implica que n puede tomar valores impares. Por lo tanto, no se puede elegir para la demostración una palabra w compuesta por n/2 símbolos b (si n es impar, ¿cuántos símbolos b constituyen la palabra?).
3. Si p es la constante del lema de bombeo elegida y p=2n, entonces la palabra w no cumple las restricciones del lema en cuanto a su longitud. Fíjate que $|w| = n + n/2 < p$ (debe ser mayor que la constante del lema de bombeo).
4. Si p es la constante del lema, ¿qué es n? Y la respuesta no puede ser que n es "la n del enunciado del problema" puesto que el problema podría enunciarse sin esta n (p.ej. diciendo "sea L el lenguaje de las palabras que tienen un número doble de aes que de bes, puden ser 0, y donde todas las aes aparecen antes de la primera b).
5. Además hay un problema con la frase "*Tomaré como constante del lema de bombeo p=2n*" que indica que probablemente no se ha entendido lo que se está haciendo y es que para demostrar que un lenguaje no es regular, no se puede elegir (tomar) la constante que queramos para el lema de bombeo, si no que hay que hacer

la demostración desconociendo esta constante (y por tanto, dándole un nombre y trabajando con él).

- **Ejemplo número 2:**

$\boxed{1.}$

Comprobamos si L es regular por medio del lema de bombeo.
Suponemos que L es regular. Entonces existe una constante n tal que $\forall w \in L$, $|w| \geq n$, $w = xyz$, $y \neq \varepsilon$, $|xy| <= n$, *teniendo que* $xy^iz \in L$.
Se elige $w = a^{2n}b^n$, $w \in L$ *y* $|w| = 3n$ *y por tanto* $|w| \geq n$.

$\boxed{2.}$

Probamos, por ejemplo, con la siguiente descomposición

- $u = a$
- $v = a$
- $x = a^{2n-2}b^n$

$\boxed{3.}$

Bombeamos:

$uv^2x = a\ a\ a\ a^{2n-2}b^n = a^{2n+1}b^n$

Podemos observar que uv^2x *no pertenece a L. Por lo tanto, el lenguaje es no regular.*

Error:

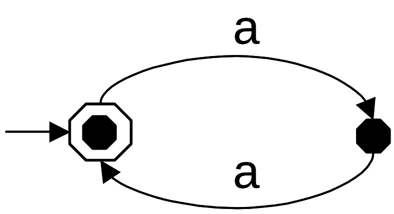
1. Se ha elegido una descomposición concreta de la palabra w, lo cual no implica que no existan otras descomposiciones válidas que puedan bombearse y, por lo tanto, el lenguaje sea regular. Recordad que para demostrar que no es regular debe cumplirse que $\forall$ x, y ,z $\in \Sigma^*$ / w = xyz, y $\neq \varepsilon$, |xy| <= n, $\exists$ k $\geq$ 0 / $xy^kz \notin$ L. Este error es cometido con frecuencia por los alumnos porque en el paso $\boxed{1.}$ de la demostración no diferencian claramente entre el enunciado del lema y lo que deben demostrar para verificar que éste no se cumple. Si se estuviera tratando de demostrar que el lema de bombeo se cumple para ese lenguaje, entonces si que se podrían elegir u, v y x, pero aquí se está tratando de demostrar lo contrario, así que hay que trabajar con cada u, v y x posibles que cumplan las restricciones del lema.

- **Ejemplo número 3:**

Esta respuesta no utiliza el lema de bombeo, pero ilustra un error típico para tratar de demostrar que un lenguaje es regular, y por tanto se ha considerado relevante incluirla aquí.

Para demostrar que es un lenguaje regular se van a aplicar las propiedades de cierre de los lenguajes regulares.
Podemos descomponer el lenguaje L tal que $L = L_1 \cdot L_2$, *siendo* $L_1 = \{ a^{2n} \mid n \geq 0\}$ *y* $L_2 = \{ b^n \mid n \geq 0\}$.

- L_1 *es un lenguaje regular porque se puede construir un AFD formado por dos estados que lo reconozca (ver la figura).*



- *$L_2 = b^*$ y, por lo tanto, también es un lenguaje regular.*

El lenguaje resultante de la concatenación de dos lenguajes regulares es también regular. Por lo tanto, L es un lenguaje regular.

Error:

1. $L \neq L_1 \cdot L_2$. L está formado por todas aquellas palabras cuyo número de símbolos a es el doble de símbolos b (notad que el índice n determina la relación entre los símbolos a y b de las palabras de L). Sin embargo, $L_1 \cdot L_2$ es el lenguaje formado por las palabra que tienen un número par de símbolos a y cualquier número de símbolos b. El error está en considerar que a^{2n} y b^n son dos partes independientes de las palabras de L.