Linguagens Não-Regulares

Area de Conhecimento em Algoritmos e Teoria - DCC/UFMG

Fundamentos de Teoria da Computacão

2021/1

Linguagens Não-Regulares: Introdução

- Os autômatos finitos podem reconhecer muitas linguagens uteis.
- Porem, estas maquinas não podem reconhecer todas as linguagens.

  Ou seja, existem linguagens não-regulares, que estão alem do poder de reconhecimento dos autômatos finitos.
- A intuição por detras da limitação dos autômatos finitos e que eles têm memoria muito limitada e, portanto, \se perdem em contagens longas".

  Intuitivamente, qualquer linguagem cujas cadeias precisem satisfazer uma condição de contagem de símbolos complicada o suficiente não pode ser uma linguagem regular.
- Aqui vamos estudar **linguagens não-regulares**, usando a poderosa ferramenta do **Lema do Bombeamento**.

O Lema do Bombeamento para linguagens regulares

Lema do Bombeamento: Definição formal

- <u>**Lema**</u> **Lema do Bombeamento (LB).** Se A e uma linguagem regular, então existe um numero p (o **comprimento de bombeamento**) tal que, se s e qualquer cadeia de A de comprimento no m nimo p, então s pode ser dividida em três partes,

$$s = xyz,$$

satisfazendo as seguintes condições:

1. para cada $i \geq 0$, $xy^iz \in A$,
2. $|y| > 0$, e
3. $|xy| \leq p$.

Demonstração. A demonstração completa e dada no livro-texto; aqui veremos sua essência.

Seja $M = (Q, \quad, \delta, q_1, F)$ um AFD que reconhece A.

Atribu mos ao comprimento de bombeamento p o total $|Q|$ de estados de M.

Primeiro note que se a linguagem não tem nenhuma cadeia com mais do que p s mbolos, o lema vale por vacuidade.

Lema do Bombeamento: Definição formal

- **Demonstração.** (Continuacão)

Caso contrario, para qualquer cadeia $s \in A$ com mais que p s mbolos, pelo menos um mesmo estado do autômato tem que ser visitado mais de uma vez. Isto acontece porque num AFD a cada s mbolo lido um estado e visitado, e se ha mais s mbolos que estados o *Princípio da Casa dos Pombos* garante que pelo menos um estado e visitado mais de uma vez.

Para ilustrar essa ideia, note que na gura abaixo o estado q_9 e visitado duas vezes ao se reconhecer a cadeia s que tem mais que $p = |Q|$ s mbolos.

$$
s = \begin{array}{ccccccccccccccccc}
 & s_1 & & s_2 & & s_3 & & s_4 & & s_5 & & s_6 & & \cdots & & s_n & \\
\uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow & & \uparrow \\
q_1 & & q_3 & & q_{20} & & q_9 & & q_{17} & & q_9 & & q_6 & & q_{35} & & q_{13}
\end{array}
$$

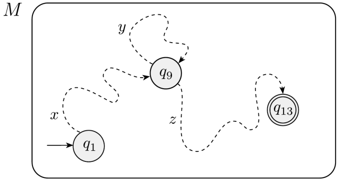
Isso quer dizer que parte da cadeia s e reconhecida em um *loop* no AFD: a subcadeia visitada entre os estados que se repetem.

No exemplo acima, os s mbolos s_4 e s_5 são consumidos dentro do *loop*.

Lema do Bombeamento: Definição formal

- **Demonstração.** (Continuacão)

  Isso quer dizer que a cadeia s pode ser quebrada em três partes $s = xyz$, em que y representa uma subcadeia não-vazia que e garantidamente consumida em um *loop*.

  Note que o fato de y ser não-vazia se re ete na Condicão (2) do lema:

  $$|y| > 0.$$

  (Mas note que ja quanto as partes x e z não temos nenhuma garantia: elas podem ou não ser vazias, e podem ou não ter *loops* tambem.)

Lema do Bombeamento: Definição formal

- **Demonstração.** (Continuação)

Note tambem que durante o reconhecimento da cadeia s, o *loop* y deve ser atingido e percorrido antes de ser ultrapassado o numero de estados do autômato.

Isto se re ete na Condição (3) do lema:

$$|xy| \leq p.$$

Logo, se o AFD M reconhece $s = xyz$, em que y e consumida em um *loop*, qualquer cadeia em que o mesmo *loop* de y seja usado um numero qualquer de vezes tambem deve ser reconhecida por M (e, portanto, tambem pertence a linguagem A).

Isto se re ete na Condição (1) do lema:

$$\forall i \geq 0 : xy^i z \in A.$$

E assim mostramos que todas as Condições (1), (2) e (3) do lema são satisfeitas para toda linguagem regular.

Lema do Bombeamento: Intuição

- Intuitivamente, o Lema do Bombeamento diz que AFDs \se perdem em contagens longas":

    1. Para qualquer cadeia grande o suficiente, o único jeito de o AFD aceitar a cadeia é fazendo pelo menos *loop*.
    2. Se um AFD reconhece uma cadeia passando por um *loop* uma vez, qualquer cadeia que passe pelo *loop* um número qualquer de vezes (e termine no mesmo estado final) também tem que ser aceita pelo AFD.

- Essa propriedade dos AFDs diz, intuitivamente, que linguagens regulares não podem ter cadeias que necessitem de contagem complexa de s mbolos.

Observações sobre o uso do Lema do Bombeamento

- Podemos usar o Lema do Bombeamento para mostrar, por contradicão, que uma linguagem L não e regular:

  1. Assumimos que L seja regular.
  2. Demonstramos que L não satisfaz o LB.
  3. Concluímos que L não pode ser regular.

- O uso do LB pode parecer dif cil de entender a princ pio; as observacões seguintes tentam explicar seu uso.

Observações sobre o uso do Lema do Bombeamento

- **Observação 1 sobre o LB:**

  O LB diz que *"Se uma linguagem L é regular, então existe uma constante p tal que para toda cadeia $s \in L$ tal que $|s| \geq p$, existe uma divisão da cadeia em $s = xyz$ para a qual as Condições (1), (2) e (3) do LB valem."*

  - Isto pode ser formalizado em lógica de predicados como:

    "Se uma linguagem L e regular, entao

    $$\exists p \in \mathbb{N} : \forall s \in L : [\ (|s| \geq p) \rightarrow (\exists\, div(s) : cond_1 \wedge cond_2 \wedge cond_3)\ ]'' , \quad (\star)$$

    onde:

    - $div(s)$ é uma função que divide a cadeia s em uma concatenação xyz;
    - $cond_1$, $cond_2$ e $cond_3$ são as Condições (1), (2) e (3) do LB, respectivamente.

Observações sobre o uso do Lema do Bombeamento

- **Observação 2 sobre o LB:**

  A nossa demonstracão por contradicão parte do princ pio de que a linguagem L e regular, mas nega a clausula $(\star)$ de conclusão do LB.

  Relembrando suas aulas de logica, note que a negacão da clausula $(\star)$ e:

  $$\forall p \in \mathbb{N} : \exists s \in L : [\ (|s| \geq p) \wedge (\forall\, div(s) : (cond_2 \wedge cond_3) \rightarrow \neg cond_1)\ ]\,.$$

  Ou seja, nossa demonstracão por contradicão

  1. assume que L seja regular,
  2. e demonstra que o LB falha para L ao mostrar que

     "para toda constante p, existe uma cadeia $s \in L$ tal que $|s| \geq p$ e, para qualquer divisão de s em xyz, se as Condições (2) e (3) do LB forem verdadeiras, a Condição (1) tem que necessariamente ser falsa."

  Como o LB não pode falhar, a unica conclusão da nossa demonstracão e que L não pode ser regular!

Lema do Bombeamento: Como usar

- Baseados nas observacões anteriores, podemos, usar o Lema do Bombeamento para mostrar, por contradicão, que uma linguagem L não e regular da seguinte forma:

  1. Assuma que L seja regular.
  2. Pelo LB, o passo (1) acima garante que existe um comprimento p tal que toda cadeia $s \in L$ com comprimento maior que p possa ser bombeada.
  3. Agora, encontre uma cadeia $s \in L$ com comprimento maior que p que não possa ser bombeada, mostrando que para <u>qualquer divisão</u> de $s = xyz$ satisfazendo

     - a Condição (2) do LB: $|y| > 0$, e
     - a Condição (3) do LB: $|xy| \leq p$,

     existe pelo menos um valor de i tal que

     - $xy^iz \notin L$, o que fere a Condição (1) do LB.
  4. Como chegamos a uma contradição, pois o passo (3) contradiz o passo (2), a hipótese do passo (1) deve ser falsa, e L não pode ser uma linguagem regular.

Lema do Bombeamento: Exemplos

- $\boxed{\text{Exemplo 1}}$ Use o Lema do Bombeamento para mostrar que a linguagem $B = \{0^n 1^n \mid n \geq 0\}$ não e regular.

Demonstração.

Vamos demonstrar por contradicão. Assuma que B seja regular, e que, portanto, ela satisfaca o Lema do Bombeamento.

Seja p o comprimento de bombeamento. Escolha s como a cadeia $0^p 1^p \in B$. Como $s \in B$ e $|s| \geq p$, o LB diz que se pode dividir s em três partes $s = xyz$, onde para qualquer $i \geq 0$, $xy^i z \in B$. Vamos considerar três casos que mostram que qualquer divisão de s torna este resultado imposs vel.

1. A cadeia y tem apenas 0s. Neste caso a cadeia $xyyz$ tem mais 0s que 1s, e não pode pertencer a B, o que contradiz a Condição (1) do LB.
2. A cadeia y tem apenas 1s. Neste caso a cadeia $xyyz$ tem mais 1s que 0s, e não pode pertencer a B, o que contradiz a Condição (1) do LB.
3. A cadeia y tem tanto 0s quanto 1s. Neste caso a cadeia $xyyz$ pode até ter o mesmo número de 0s e 1s, mas alguns 1s vêm antes de alguns 0s. Logo $xyyz$ não pode pertencer a B, o que contradiz a Condição (1) do LB. $\square$

Lema do Bombeamento: Exemplos

- $\boxed{\text{Exemplo 2}}$ Use o Lema do Bombeamento para mostrar que a linguagem $C = \{w \in \{0,1\}^* \mid w$ tem o mesmo numero de 0s e 1s$\}$ nao e regular.

Demonstração.
Vamos demonstrar por contradicao. Assuma que C seja regular, e que, portanto, ela satisfaca o Lema do Bombeamento.

Seja p o comprimento de bombeamento. Escolha s como a cadeia $0^p1^p \in C$. Como $s \in C$ e $|s| \geq p$, o LB diz que se pode dividir s em tres partes $s = xyz$, onde para qualquer $i \geq 0$, $xy^iz \in C$.

Nesta demonstracao vamos usar a Condicao (3) do LB, que diz que $|xy| \leq p$. Como s comeca com 0^p, isto quer dizer que y so pode ter 0s. Alem disso, a Condicao (2) diz que y tem pelo menos um 0.

Neste caso a cadeia $xyyz$ tem mais 0s que 1s, e nao pode pertencer a C, o que contradiz a Condicao (1) do LB. □

- $\boxed{\text{Exemplo 3}}$ Use o Lema do Bombeamento para mostrar que a linguagem $F = \{ww \mid w \in \{0,1\}^*\}$ não e regular.

  Demonstração.
  Suponha, ao contrario, que F seja regular. Seja p o comprimento de bombeamento dado pelo LB. Seja s a cadeia $0^p1^p0^p1^p \in F$. Como $s \in F$ e $|s| \geq p$, o LB garante que s pode ser dividida em três partes $s = xyz$ satisfazendo as três condicões do lema. Mostramos que isso não e poss vel.

  Vamos novamente usar a Condicão (3) do LB, que diz que $|xy| \leq p$. Isso quer dizer que y so pode ter 0s, uma vez que s comeca com p 0s. Logo, a cadeia $xyyz$ sera da forma $0^k1^p0^p1^p$, com $k \geq p$, e, portanto não pertence a L. Isto contradiz o LB. □

Lema do Bombeamento: Exemplos

- $\boxed{\text{Exemplo 4}}$ Use o Lema do Bombeamento para mostrar que a linguagem $D = \{1^{n^2} \mid n \geq 0\}$ não e regular.

Demonstração.
Suponha, ao contrario, que D seja regular. Seja p o comprimento de bombeamento dado pelo LB. Seja s a cadeia $1^{p^2} \in D$. Como $s \in D$ e $|s| \geq p$, o LB garante que s pode ser dividida em três partes $s = xyz$ satisfazendo as três condicões do lema. Mostramos que isso não e poss vel.

Para esta demonstracão, temos que ter um tanto especial de criatividade.

Note que a sequência de quadrados perfeitos e

$$0, 1, 4, 9, 16, 25, 36, 40, \ldots,$$

de forma que a distância entre dois membros sucessivos da sequência e crescente.

- $\boxed{\text{Exemplo 4}}$ (Continuacão)

Agora considere as duas cadeias xyz e xy^2z, cujos tamanhos diferem pelo comprimento da subcadeia y a mais na segunda cadeia.

Note que pela condicão (3) do LB, $|xy| \leq p$, logo $|y| \leq p$.

Como temos que $|xyz| = p^2$, então $|xy^2z| \leq p^2 + p$.

Mas $p^2 + p < p^2 + 2p + 1 = (p + 1)^2$. Como a condicão (2) diz que $|y| > 0$, temos que $|xy^2z| > p^2$.

Juntando nossas conclusões, temos que $p^2 < |xy^2z| < (p + 1)^2$.

Ou seja, $|xy^2z|$ esta estritamente entre dois quadrados perfeitos consecutivos, e, portanto, $|xy^2z|$ não pode ser um quadrado perfeito.

Dessa forma, chegamos a contradicão de que $xy^2z \neq D$ e conclu mos que D não e regular.

$\square$

Lema do Bombeamento: Exemplos

- $\boxed{\text{Exemplo 5}}$ Use o Lema do Bombeamento para mostrar que a linguagem $E = \{0^i 1^j \mid i > j\}$ não é regular.

Demonstração.
Suponha, ao contrario, que E seja regular. Seja p o comprimento de bombeamento dado pelo LB. Seja s a cadeia $0^{p+1}1^p \in E$. Como $s \in E$ e $|s| \geq p$, o LB garante que s pode ser dividida em três partes $s = xyz$ satisfazendo as três condicões do lema. Mostramos que isso não é possível.

Pela Condicão (3) do LB, y contem somente 0s, e pela condicão (2) y tem pelo menos um 0. Logo a cadeia $xy^0z = xz$ é da forma $0^k 1^p$, com $k \leq p$ (pois os 0s de y foram retirados da cadeia), o que contradiz a Condicão (1) do lema.

□

Observações sobre o uso do Lema do Bombeamento

- Note que o LB diz que se uma linguagem e regular, então a linguagem satisfaz uma serie de condições.

  Por isso, o LB serve <u>APENAS</u> para mostrar que
  <u>uma linguagem NÃO e regular</u>.

- Entretanto, o LB não e um \se, e somente se": sua conversa não e verdadeira.

  Isto e, existem linguagens que não são regulares, mas que satisfazem a serie de condições do LB, como as seguintes:

  1. $\{\mathtt{ab}^n\mathtt{c}^n \mid n \geq 0\} \cup \{\mathtt{a}^k w \mid k \geq 2,\ w \in \Sigma^* \text{ não começa com } \mathtt{a}\}$
  2. $\{ww^R x \mid w, x \in \{0, 1\}^+\}$

  (Um desa o para você e mostrar que tais linguagens: (i) satisfazem o LB; mas (ii) não são regulares.)

  Por isso, o LB <u>NÃO</u> pode ser usado para mostrar que
  <u>uma linguagem E regular</u>.

Provando que uma linguagem não é regular usando propriedades de fechamento

Observações sobre o uso do Lema do Bombeamento

- Ja vimos que as linguagens regulares são fechadas sob as operacões de:
  - união,
  - interseção,
  - concatenação,
  - fecho de Kleene, e
  - complemento.

- Vimos que podemos demonstrar que uma linguagem e regular usando propriedades de fechamento.

  Mais especi camente, vimos que podemos demonstrar que uma L e regular ao escrever L como a composicão de linguagens regulares usando as operacões acima.

- Entretanto, podemos usar as propriedades de fechamento para demonstrar que uma linguagem <u>não</u> e regular.

- Exemplo 6 Usando o Lema do Bombeamento, ja demonstramos que a linguagem $C = \{w \in \{0,1\}^* \mid w$ tem o mesmo numero de 0s e 1s$\}$ não e regular. Mostre que a linguagem C não e regular usando a propriedade de fechamento sobre linguagens regulares.

Demonstração.
Vamos comecar notando que a linguagem 0^*1^* e regular, pois ela e representada por uma expressão regular.

Agora, note que $C \cap 0^*1^* = \{0^n1^n \mid n \geq 0\}$, pois C e a linguagem em que cada cadeia tem o mesmo numero de 0s e 1s, e $\{0^n1^n \mid n \geq 0\}$ e a linguagem em que todos os 1s vêm apos todos os 0s.

Mas como as linguagens regulares são fechadas sob intersecão, se C fosse regular, sua intersecão com a linguagem regular 0^*1^* tambem deveria ser regular, mas não e: $\{0^n1^n \mid n \geq 0\}$ não e regular.

Logo, conclu mos que C não pode ser regular. □