Linguagens Não-Regulares

Para entender o poder dos autômatos finitos temos também que entender suas limitações. Mostraremos como provar que certas linguagens não podem ser reconhecidas por nenhum autômato finito.

Tomemos a linguagem $B = \{0^n 1^n \mid n \geq 0\}$.

Se tentarmos encontrar um AFD que reconheça B, descobriremos que a máquina parece necessitar de lembrar quantos 0's foram vistos até então à medida que ela lê a entrada.

Como o número de 0's não é limitado, a máquina terá de registrar um número ilimitado de possibilidades. Mas ela não pode fazer isso com qualquer quantidade finita de estados.

Um método para provar que linguagens como a B não são regulares.

O argumento que acaba de ser apresentado já não prova não-regularidade baseando-se no fato de que o número de 0's é ilimitado? Não, não prova.

Só porque a linguagem parece requerer memória ilimitada não significa que assim seja necessariamente. Isso acontece de ser verdadeiro para a linguagem B, mas outras linguagens parecem requerer um número ilimitado de possibilidades, e ainda assim, são regulares.

Exemplos:
Considere as duas linguagens sobre o alfabeto $\Sigma = \{0, 1\}$:

$C = \{w \mid w$ tem um número igual de 0s e 1s$\}$.

$D = \{w \mid w$ tem um número igual de ocorrências de 01 e 10 como subcadeias$\}$.

À primeira vista, uma máquina reconhecedora parece ter que contar em ambos os casos, e portanto nenhuma das duas parece ser regular. Como esperado, C não é regular, mas surpreendentemente D é regular.

O Lema do Bombeamento para Linguagens Regulares

A técnica para provar não-regularidade provém do teorema sobre linguagens regulares, chamado de lema do bombeamento.

Esse teorema afirma que todas as linguagens regulares têm uma propriedade especial.
Se pudermos mostrar que uma linguagem não tem essa propriedade, temos a garantia de que ela não é regular.

A propriedade enuncia que todas as cadeias da linguagem podem ser "bombeadas" se elas são no mínimo tão longas como um determinado valor especial, denominado o comprimento de bombeamento.
Isso significa que cada uma dessas cadeias contém uma parte que pode ser repetida um número qualquer de vezes, com a cadeia resultante

permanecendo na linguagem.

Teorema 1.70:

Lema do Bombeamento: Se A é uma linguagem regular, então existe um número p (o comprimento de bombeamento) tal que, se s é qualquer cadeia de A de comprimento no mínimo p, então s pode ser dividida em três partes, $s = xyz$, satisfazendo as condições:

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

A notação $|s|$ representa o comprimento da cadeia s, y^i significa que i cópias de y são concatenadas entre si e y^0 é igual a ε.

Quando s é dividida em xyz, ou x ou z pode ser ε, mas a condição 2 diz que $y \neq \varepsilon$. Observe que sem a condição 2 o teorema seria trivialmente verdadeiro.

A condição 3 enuncia que as partes x e y juntas têm comprimento máximo p.
Ela é uma condição técnica extra que ocasionalmente achamos útil ao provar que certas linguagens não são regulares.

Idéia de prova:

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

Atribuímos ao comprimento de bombeamento p o número de estados de M.

Mostramos que qualquer cadeia s em A de comprimento pelo menos p pode ser quebrada em três partes xyz satisfazendo as nossa três condições.

E se nenhuma cadeia de A tem comprimento no mínimo p?

Então nossa tarefa é ainda mais fácil, visto que o teorema se torna verdadeiro por vacuidade: obviamente, as três condições se verificam para todas as cadeias de comprimento no mínimo p se não existem tais cadeias.

Se s em A tem comprimento pelo menos p, considere a sequência de estados pelos quais M passa quando computa com a entrada s.

Ele começa no estado ~~q_1~~ inicial, q_1, vai para, digamos, q_3, depois para, digamos, q_{20}, e em seguida para ~~q~~ q_9, e assim por diante, até que atinge o final de s no estado q_{13}.

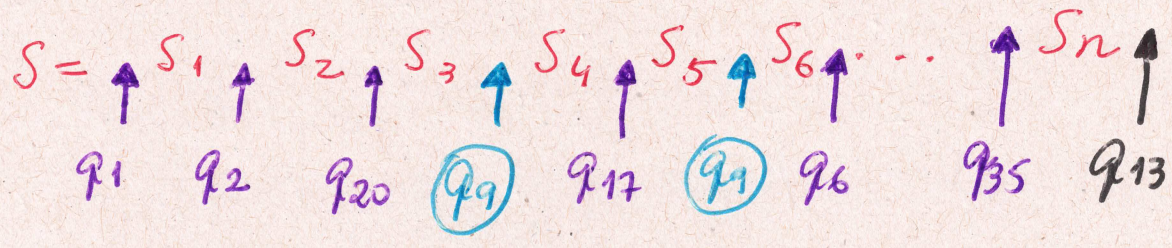
Com s em A, sabemos que M aceita s, portanto q_{13} é um estado de aceitação.

Se dissermos que n é o comprimento de s, a seqüência de estados $q_1, q_3, q_{20}, q_9, \ldots, q_{13}$ tem comprimento $n+1$.

Como n é no mínimo p, sabemos que $n+1$ é maior que p, o número de estados de M.
Assim, a seqüência tem de conter um estado repetido.

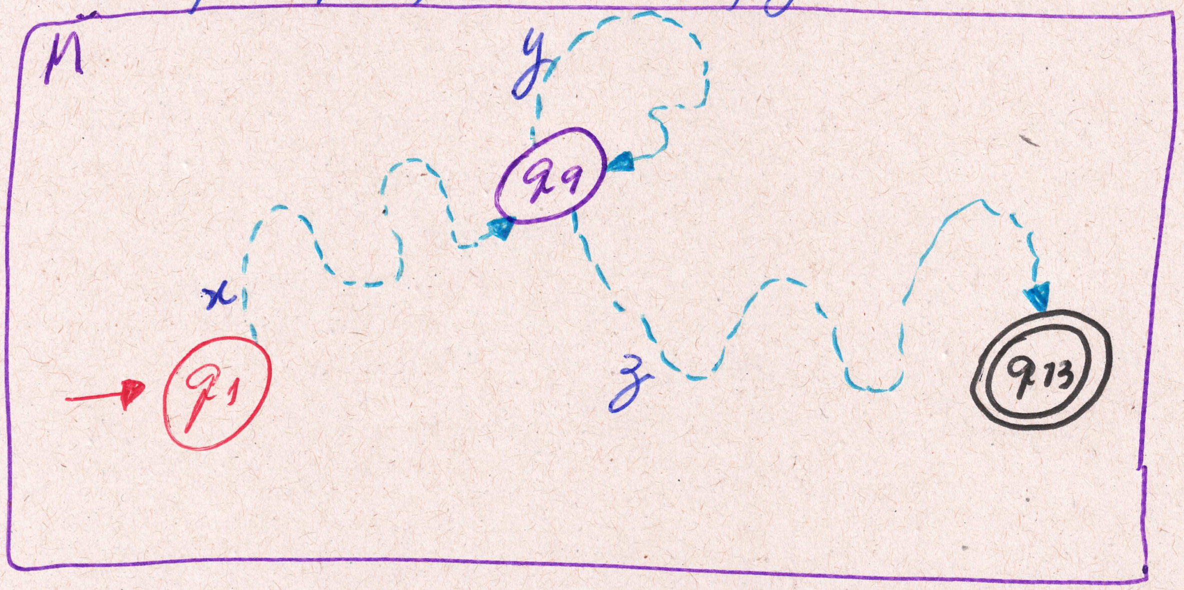
Esse resultado é um exemplo do princípio da casa de pombos, "se p pombos forem colocados em menos que p casas, alguma casa recebe mais de um pombo".

A figura 1.71 mostra a cadeia s e a seqüência de estados pelos quais M passa quando processa s. O estado q_9 é o que se repete.



Agora dividimos s nas três partes x, y e z.
A parte x é a parte de s que aparece antes de q_9, a parte y é a que ocorre entre as duas aparições de q_9, e a parte z é o resto de s, vindo após a segunda ocorrência de q_9. Portanto, x leva M do estado q_1 para q_9, y leva M de q_9 de volta para q_9 e z leva M de q_9 para o estado de

aceitação q_{13}, como a figura 1.72.



Vamos ver que por essa divisão de s satisfaz as três condições.

Suponha que rodemos M sobre a entrada $xyyz$. Sabemos que x leva M de q_1 para q_9, e então o primeiro y o leva de q_9 de volta para q_9, assim como faz o segundo y, e depois z o leva para q_{13}.

Com q_{13} sendo um estado de aceitação, M aceita a entrada $xyyz$.

Analogamente, ele irá aceitar xy^iz para qualquer $i>0$. Para o caso $\boxed{i=0}$, $xy^iz = xz$, que é aceita por razões semelhantes. Isso estabelece a condição 1.

Verificando a condição 2, vemos que $|y|>0$, pois y era a parte de s que apareceria entre duas ocorrências diferentes do estado q_9.

Para assegurar a condição 3, façamos com que q_9 seja a primeira repetição na seqüência.

Pelo princípio da casa de pombros, os primeiros $p+1$ estados na seqüência devem conter uma repetição.

Por conseguinte, $|xy| \leq p$.

Prova: Seja $M = (Q, \Sigma, \delta, q_1, F)$ um AFD que reconhece A e p o número de estados de M.

Seja $s = s_1 s_2 \dots s_n$ uma cadeia em A de comprimento n, onde $n \geq p$. Seja $r_1, \dots, r_{n+1}$ a seqüência de estados nos quais M passa enquanto processa s, de forma que $r_{i+1} = \delta(r_i, s_i)$ para $1 \leq i \leq n$. Essa seqüência tem comprimento $n+1$, que é pelo menos $p+1$. Entre os primeiros $p+1$ elementos da seqüência, dois devem ser o mesmo estado, pelo princípio da casa de pombos.

Chamamos o primeiro desses de r_j e o segundo de r_l. Como r_l ocorre entre as primeiras $p+1$ posições da sequência começando em r_1, temos que $l \leq p+1$. Agora, seja $x = s_1 \dots s_{j-1}$, $y = s_j \dots s_{l-1}$ e $z = s_l \dots s_n$.

Como x leva M de r_1 para r_j, y leva M de r_j para r_j e z leva M de r_j para r_{n+1}, que é um estado de aceitação, M deve aceitar xy^iz para $i \geq 0$. Sabemos que $j \neq l$, e portanto $|y| > 0$; e $l \leq p+1$, e logo $|xy| \leq p$. Dessa forma, satisfazemos todas as condições do lema de bombeamento.

Para usar o lema de bombeamento para provar que uma linguagem B não é regular, primeiro suponha que B seja regular, a fim de obter uma contradição.

Então use o lema do bombeamento para garantir a existência de um comprimento de bombeamento p de forma que todas as cadeias de comprimento p ou maiores em B possam ser bombeadas.

Em seguida, encontre uma cadeia s em B que tenha comprimento p ou mais, mas que não possa ser bombeada.

Finalmente, demonstre que s não pode ser bombeada considerando todas as maneiras de dividir s em x, y e z (levando a condição 3 do lema em consideração se for conveniente) e, para cada divisão, encontrando um valor i tal que $xy^iz \notin B$.

Esse passo final freqüentemente envolve agrupar as várias formas de se dividir s em vários casos e analisá-los individualmente. A existência de s contradiz o lema do bombeamento se B for regular. Assim, B não pode ser regular.

Exemplo 1.73:

Seja B a linguagem $\{0^n 1^n \mid n \geq 0\}$.
Usamos o lema de bombeamento para provar que B não é regular. A prova é por contradição.

Suponha, ao contrário, que B seja regular.
Seja p o comprimento de bombeamento dado pelo lema do bombeamento. Escolha s como a cadeia $0^p 1^p$. Como s é um membro de B e tem comprimento maior que p, o lema do bombeamento garante que s pode ser dividida em três partes, $s = xyz$, onde para qualquer $i \geq 0$ a cadeia xy^iz está em B. Consideramos três casos para mostrar que esse resultado é impossível.

1. A cadeia y contém apenas 0s. Neste caso, a cadeia $xyyz$ tem mais 0s que 1s e, portanto, não é um membro de B, violando a condição 1 do lema de bombeamento. Esse caso é uma contradição.

2. A cadeia y contém somente 1s. Esse caso também é uma contradição.

3. A cadeia y contém ambos, 0s e 1s. Nesse caso, a cadeia $xyyz$ pode ter o mesmo número de 0s e 1s, mas eles estarão fora de ordem, com alguns 1s antes de 0s. Logo, ela não é um membro de B, o que é uma contradição.

Assim, uma contradição é inevitável se fizermos a suposição de que B é regular; portanto, B não é regular.

Note que podemos simplificar esse argumento aplicando a condição 3 do lema para eliminar os casos 2 e 3.

Nesse exemplo, encontrar a cadeia s foi fácil, porque qualquer cadeia ~~de~~ em B de comprimento p ou mais funcionaria.

Exemplo 1.74:

Seja $C = \{w \mid w$ tem número igual 0s e 1s$\}$.

Usamos o lema do bombeamento para provar que C não é regular. ~~A por~~

A prova é por contradição. [pág 83, Sipser].

Suponha, ao contrário, que C seja regular.

Seja p o comprimento de bombeamento dado pelo lema.

Seja s a cadeia 0^p1^p. Como sendo s um membro de C e tendo comprimento maior que p, o lema do bombeamento garante que s pode ser dividida em três partes, $s = xyz$, onde para qualquer $i \geq 0$ a cadeia xy^iz está em C.

Gostaríamos de demonstrar que isso é impossível.

Mas espere, isso é possível!

Se fizermos x e z serem a cadeia vazia e y ser a cadeia 0^p1^p, então xy^iz sempre terá um número igual de 0s e 1s e portanto, está em C.

Logo, parece que s pode ser bombeada.

Aqui a condição 3 no lema é útil. Ela estipula que, quando bombeando s, ela deve ser dividida de forma que $|xy| \leq p$. Essa restrição na maneira em que s pode ser dividida torna mais fácil mostrar que a cadeia $s = 0^p1^p$ que selecionamos não pode ser bombeada.

12

Se $|xy| \leq p$ então y deve conter somente 0s, logo $xyyz \notin C$. Por conseguinte, s não pode ser bombeada. O que nos dá a contradição desejada.

A seleção da cadeia s nesse exemplo exigiu mais cuidado do que no Exemplo 1.73. Se tivéssemos escolhido $s = (01)^p$, teríamos tido problema, pois precisamos de uma cadeia que não possa ser bombeada e essa cadeia pode, mesmo levando-se em consideração a condição 3. Uma maneira de bombeá-la seria com $x = \varepsilon$, $y = 01$ e $z = (01)^{p-1}$. Com isso, $xy^iz \in C$ para todo valor de i.