Autômatos determinísticos grandes

Arnaldo Mandel

27 de outubro de 2009

A construção dos subconjuntos implica na seguinte afirmativa: *se uma linguagem é reconhecida por um autômato não-determinístico com n estados, então ela é reconhecida por um autômato determinístico com* $\leq 2^n$ *estados.*

Há uma pequena ressalva a ser feita na afirmativa acima, devido à variedade de definições de autômatos não-determinísticos. O cuidado é que o rótulo de cada transição seja uma letra ou a palavra vazia λ. Se outras palavras são admitidas como rótulos, as contas mudam.

Ao ver esse fato, muito se perguntam quanto à sua precisão. Tudo bem que 2^n é um limitante, mas qual a limitação real? Não é difícil formular esta questão rigorosamente; basta lembrar que cada linguagem regular é reconhecida por um autômato determinístico reduzido único.

Questão 1 *Para cada inteiro n, considere as linguagens reconhecidas por autômatos com n estados, e seja* $f(n)$ *o maior número de estados para um autômato reduzido reconhecendo uma dessas linguagens. Determinar* $f(n)$*. Ou pelo menos, quanto cresce* $f(n)$*.*

Nesses termos, já sabemos que $f(n) \leq 2^n$, assim, o importante é determinar um limitante inferior. A resposta é conhecida há muito tempo, e este texto é uma pequena exposição.

Um primeiro exemplo já indica por onde vão as coisas.

Exemplo 1: Para cada n, seja $\Sigma_n = \{1, 2, \ldots, n\}$, e seja $L_n = \{x \in \Sigma_n^* : x$ não contêm todas as letras do alfabeto$\}$. Uma expressão regular: $L_n = \bigcup_i (\Sigma_n - \{i\})^*$. Essa expressão leva facilmente ao autômato não-determinístico $\mathcal{A}_n = (K, \Sigma_n, \Delta, K, K)$, onde $K = \Sigma_n$, e Δ consiste de todas as triplas (i, j, i) com $j \neq i$. ▬

Proposição 1 *O autômato reduzido para* L_n *tem* 2^n *estados.*

Prova: A construção dos subconjuntos para $\mathcal{A}_n$ nos dá o autômato determinístico $\mathcal{D}_n = (2^K, \Sigma_n, \delta, K, \mathcal{F})$, onde $\mathcal{F}$ consiste de todos os subconjuntos não vazios de K, e $\delta(X, i) = X - i$, para cada $X \subseteq K$, $i \in \Sigma_n$. Vamos mostrar que esse autômato é reduzido.

Começamos com a seguinte propriedade: para cada $X \subseteq K$ e toda palavra w, $\delta(X, w) = X - \alpha(w)$, onde $\alpha(w)$ é o conjunto de letras que ocorrem em w. Isso segue trivialmente por indução em $|w|$, e fica de exercício.

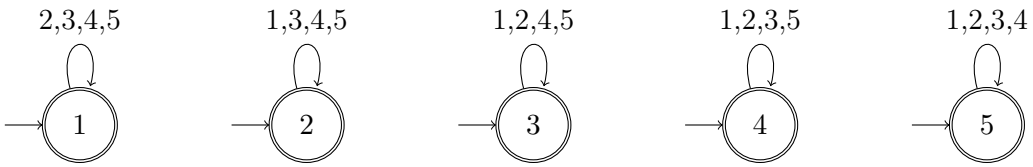

Figura 1: O autômato $\mathcal{A}_5$

Agora, vamos mostrar:

a) Todos estados são acessíveis: para cada subconjunto $X \subseteq K$, seja w uma palavra contendo precisamente as letras que não estão em X. Segue da propriedade acima que $\delta(K, w) = X$.

b) Todos estados são não equivalentes entre si: dados dois estados X, Y, sem perda de generalidade podemos supor que $Y - X \neq \emptyset$; seja w uma palavra obtida multiplicando-se todas as letras de Y em alguma ordem. Então, $\delta(X, w) = \emptyset \neq \delta(Y, w)$.

□

Isto mostra que $f(n) = 2^n$, mas deixa uma certa insatisfação. Afinal das contas, esses autômatos são sobre alfabetos progressivamente maiores. O que acontece se mantivermos um alfabeto fixo?

Questão 2 *Fixe um alfabeto Σ. Para cada inteiro n, considere as linguagens reconhecidas por autômatos com n estados, e seja $f(n)$ o maior número de estados para um autômato reduzido reconhecendo uma dessas linguagens. Determinar $f(n)$. Ou pelo menos, quanto cresce $f(n)$.*

Começamos por um alfabeto de uma letra, a.

Exemplo 2: Sejam $n_1, n_2, \ldots, n_k$ inteiros positivos, dois a dois primos entre si, tais que $n_1 + n_2 + \cdots + n_k = n$, onde o número k de parcelas é arbitrário. O autômato $\mathcal{P}_{n_1,n_2,\ldots,n_k}$ tem como grafo uma coleção de ciclos de comprimentos $n_1, n_2, \ldots, n_k$, um estado inicial em cada ciclo e os estados finais são os iniciais. Então, veremos, o número de estados do autômato reduzido para $\mathcal{L}(\mathcal{P}_{n_1,n_2,\ldots,n_k})$ é o produto $n_1 n_2 \cdots n_k$. ▬

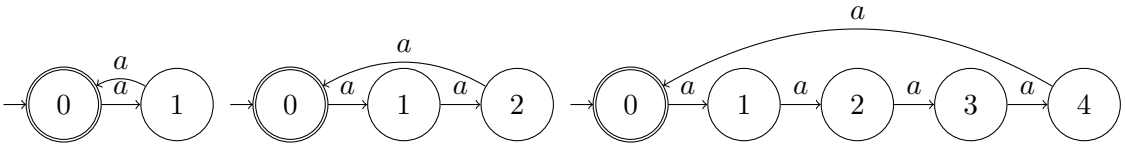

Figura 2: O autômato $\mathcal{P}_{2,3,5}$

Para conferir o número de estados, vamos numerar os estados do ciclo j usando os inteiros módulo n_j, conforme a figura. Assim, todas as arestas vão do vértice i ao vértice $i+1 \mod n_j$. Aplicando a construção dos subconjuntos, obtemos um autômato determinístico $\mathcal{D}$ cujo estado inicial S consiste dos k estados de rótulo 0. É fácil ver que para qualquer palavra a^r, $\delta(S, a^r)$ contem exatamente um estado de cada ciclo, assim, os estados acessíveis de $\mathcal{D}$ podem ser escritos como k-uplas $\mathbf{p} \in \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$, e $\delta(\mathbf{p}, a)$ simplesmente soma 1 em cada coordenada. O Teorema Chinês do Resto "diz" que todas essas k-uplas são estados acessíveis. Uma outra aplicação, um pouco mais cuidadosa, do mesmo Teorema, prova que esses estados são todos não equivalentes.

Quão grande pode ser m, dado n? Para ser mais preciso, o que queremos é estimar $g(n) = \max\{\text{mmc}(n_1, n_2, \ldots, n_k) : n_1 + n_2 + \cdots + n_k = n\}$; essa função foi estudada no início do século XX pelo russo Landau, e é conhecida na literatura pelo nome de função de Landau. Ele provou que

$$\lim_{n\to\infty} \frac{\ln g(n)}{\sqrt{n \ln n}} = 1.$$

Posteriormente foi provado que

$$g(n) = e^{\sqrt{n(\ln n + \ln\ln n - 1 + o(1))}}.$$

Uma referência mais ou menos legível para isso é
W. Miller, The Maximum Order of an Element of Finite Symmetric Group, *Am. Math. Monthly*, **94** (1987), 497–506.

Essas expressões mostram claramente que $g(n)$ cresce muito mais que qualquer polinômio, mas muito menos que 2^n. Será que dá para chegar significativamente mais perto de 2^n com uma única letra? A resposta é **não**: não dá para passar de $g(n) + n$.

Pode assustar um pouco essa situação com uma letra. Afinal, tanta conta sofisticada, o que será que acontece com duas letras? Não pode ficar tão ruim, afinal, vimos acima que se o número de letras é grande dá para controlar o resultado.

Uma idéia é extrair o suco dos $\mathcal{A}_n$:

Proposição 2 *Seja $\mathcal{A}$ um autômato não determinístico com n estados, todos eles iniciais e finais, e sem transições λ. Suponha que para cada estado p, existe uma palavra w_p tal que:*

1. *Não existe passeio com rótulo w_p iniciando em p.*
2. *Para todo estado $q \neq p$, existe um único passeio com rótulo w_p com início em q, e esse passeio termina em q.*

Então o autômato reduzido para $\mathcal{L}(\mathcal{A})$ tem 2^n estados.

Prova: Seja K o conjunto de estados de $\mathcal{A}$ e δ a função de transição do autômato $\mathcal{D}$ obtido de $\mathcal{A}$ pela construção dos subconjuntos. Note que, como não existem transições λ, todo subconjunto

de K é fechado. Como todo estado de $\mathcal{A}$ é final, o único estado não final de $\mathcal{D}$ é $\emptyset$. Pelas propriedades (1) e (2) segue que:

Para cada $X \subseteq K$ *e* $p \in K$, $\delta(X, w_p) = X - p$.

Logo:

1. Todo subconjunto X de K é acessível em $\mathcal{D}$. Isso porque, se $K - X = \{p_1, p_2, \ldots, p_k\}$, a propriedade acima implica que $\delta(K, w_{p_1} w_{p_2} \cdots w_{p_k)} = X$.

2. Subconjuntos distintos de K são não equivalentes em $\mathcal{D}$. Para ver isso, sejam $X, Y \subseteq K$, e suponha que $q \in X - Y$. Se $Y = \{p_1, p_2, \ldots, p_k\}$, então $\delta(Y, w_{p_1} w_{p_2} \cdots w_{p_k}) = \emptyset$, enquanto que $q \in \delta(K, w_{p_1} w_{p_2} \cdots w_{p_k})$.

Daí segue que $\mathcal{D}$ é o autômato reduzido para $\mathcal{L}(\mathcal{A})$. □

Agora podemos voltar ao Exemplo 1 e aplicar a proposição acima. A palavra w_i é simplesmente i. Pronto, a Proposição 1 vira um corolário trivial da Proposição 2. Nenhuma novidade aí, a demonstração é a mesma. Só que agora podemos construir uma família de exemplos mais interessantes.

EXEMPLO 3: Vamos trabalhar com o alfabeto binário, $\Sigma = \{a, b\}$. Dado o inteiro positivo n, seja $\mathcal{B}_n$ o autômato com estados $K = \{1, 2, \ldots, n\}$, todos eles iniciais e finais, e seja $\Delta = \{(i, a, i-1), (i, b, i-1) : 1 < i \leq n\} \cup \{(1, a, n)\}$. ▬

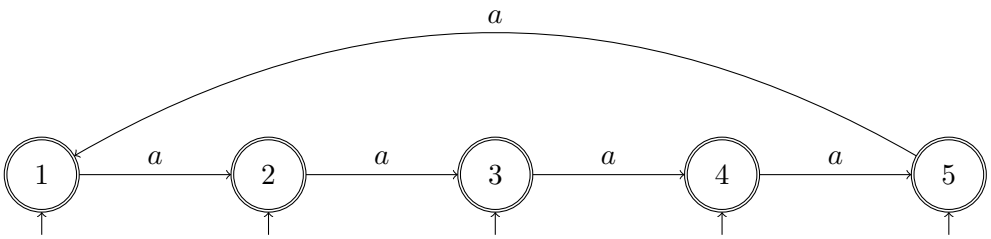

Figura 3: O autômato $\mathcal{B}_5$

Vamos verificar que $\mathcal{B}_n$ satisfaz as condições da Proposição 2. Tudo certo quanto a estados iniciais e finais, e transições λ. Para isso, é preciso dar uns w_p. Aí estão algumas palavras que funcionam:

$$w_p = a^{p-1}ba^{n-p}.$$

A partir do estado p, existe um caminho com rótulo a^{p-1}, que termina no estado 1; daí, não há aresta com b, assim, a condição (1) é satisfeita. A partir de outro estado, o caminho com rótulo a^{p-1} leva a um estado diferente de 1, daí que existe aresta com b, e o resto dos a's levam ao estado de partida; daí que (2) é satisfeito.

Pronto, temos que o autômato reduzido para $\mathcal{L}(\mathcal{B}_n)$ tem 2^n estados, e *só usamos duas letras.*

Alguns autores não gostam de autômatos com mais de um estado inicial. Eu não sei como alavancar a Proposição 2 para produzir 2^n estados no determinístico com um único estado inicial e só duas letras. Dá para chegar "meio perto": introduza mais um estado em $\mathcal{B}_{n-1}$, faça ele ser o único estado inicial, e coloque arestas de rótulo a indo desse estado para todos os outros. É fácil ver que o autômato reduzido do resultado tem $2^{n-1} + 1 > \frac{1}{2}2^n$ estados.

Agora, com três letras é bico: é só alterar $\mathcal{B}_n$, colocando arestas $(1, c, i), i = 1, 2, \ldots, n$, e fazer 1 ser o estado inicial. Esse vai ter um reduzido com 2^n estados, e é um autômato não determinístico prá ninguém botar defeito.

Uma curiosidade: qual o menor tamanho de uma expressão regular para $\mathcal{L}(\mathcal{B}_n)$? Consegui uma expressão de tamanho aproximadamente $18n + 4$ (lembrando que o comprimento conta só ocorrências de $a, b, \lambda, +$ e *; fora isso, foram $4n + 1$ pares de parênteses). Alguém consegue algo melhor?

EXEMPLO 4: Finalmente, uma seqüência de autômatos sobre duas letras e com um único estado inicial. O autômato $\mathcal{C}_n$ tem estados $K = \{0, 1, \ldots, n\}$; suas transições são $\{(i, a, i+1), (i, b, i+1) : i = 1, 2, \ldots n\} \cup \{(n, a, 0), (n, a, 1), (0, a, 1), (0, b, 0)\}$. O estado 0 inicial e final, e não há outros. ▬

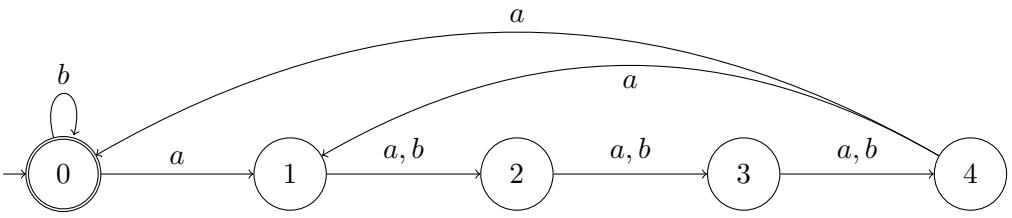

Figura 4: O autômato $\mathcal{C}_4$

Mostrar que o autômato determinístico correspondente tem 2^{n+1} estados parece bem mais difícil do que no caso anterior. Abaixo segue um roteiro de demonstração; completá-la fica como exercício. Se alguém conseguir uma demonstração mais simples, eu gostaria de ver.

Seja δ a função de transição do autômato $\mathcal{D}_n$ obtido de $\mathcal{C}_n$ pela construção dos subconjuntos.

1. Vamos mostrar que todo subconjunto de K é acessível em $\mathcal{D}_n$.
   (a) Como $\delta(\{0\}, ab^n) = \emptyset$ e $\delta(\{0\}, a^j) = \{j\}, j = 1, 2, \ldots, n$, todo conjunto com menos que dois elementos é acessível.
   (b) Represente cada subconjunto de K pela sequência ordenada de seus elementos. Agora, argumente por indução na ordem comprimento-lexicográfica; seja $X = i_1 < i_2 < \cdots < i_k$.
   CASO 1: $i_1 > 0$.
   Então $X = \delta(i_1 - 1 < i_2 - 1 < \cdots < i_k - 1, a)$.
   CASO 2: $i_1 = 0$.
   CASO 2.1: $i_2 > 1$.
   Então $X = \delta(i_1 < i_2 - 1 < \cdots < i_k - 1, b)$.
   CASO 2.2: $i_2 = 1$.
   Então $X = \delta(i_3 - 1 < \cdots < i_k - 1 < n, a)$.
2. Agora vamos mostrar que conjuntos distintos são não equivalentes em $\mathcal{D}_n$. Dados $X, Y \subseteq K$ distintos, seja k um elemento na diferença simétrica de X e Y, e sem perda de generalidade, suponha que $k \in X$. Então, se $k = 0$, claro que X e Y não são equivalentes; se $k \neq 0$, então $0 \in \delta(X, b^{n-k}a)$ e $0 \notin \delta(Y, b^{n-k}a)$.

A linguagem $\mathcal{L}(\mathcal{C}_n)$ tem uma expressão regular razoavelmente simples, de comprimento $6n$:

$$(b + a((a+b)^{n-1}a)^*(a+b)^{n-1}a)^*,$$