El Teorema de Schützenberger

Alejandro Mallea

Pontificia Universidad Católica de Chile

Recuerdo de Teoría de Autómatas

Consideramos algunos conceptos base de Teoría de Autómatas:

- alfabeto (finito),
- concatenación, palabras,
- Σ^*, lenguajes,
- expresión regular, lenguaje regular,
- autómata finito determinista.

Recuerdo de Matemáticas Discretas

Consideramos algunos conceptos base de Matemáticas Discretas:

- relación de equivalencia, clases de equivalencia,
- cuociente de un conjunto por las clases de equivalencia
- índice de una relación de equivalencia.

Ejemplo de Relación de Equivalencia

Consideremos el conjunto $X = \mathbb{Q}$ y la relación $\sim$ sobre X:

$$x \sim y \Leftrightarrow x - y \in \mathbb{Z}.$$

Clases de equivalencia: $[x]_\sim$, para $x \in X$ (o simplemente $[x]$).

Cuociente: $X/\sim \stackrel{\text{def}}{=} \{[x]_\sim : x \in X\}$.

En este caso, $X/\sim = \{[x] : 0 \leq x < 1\}$.

El índice de una relación $\sim$ es el tamaño de $X/\sim$.
La relación del ejemplo tiene índice infinito.

Lenguajes *star-free*

Consideremos expresiones regulares extendidas con un operador de complementación:

- Si r es e.r., $\overline{r}$ es e.r. extendida, y $L(\overline{r}) = \overline{L(r)} = \Sigma^* \setminus L(r)$.
- Si r_1, r_2 son e.r., $r_1 \cap r_2$ es una expresión regular extendida, y $L(r_1 \cap r_2) = L(r_1) \cap L(r_2)$.
  - $\cap$ no es estrictamente necesario: $L(r_1 \cap r_2) = L\left(\overline{\overline{r_1} + \overline{r_2}}\right)$.

Definición

Un lenguaje es star-free si se puede generar con una expresión regular extendida que no usa el operador *.

Ejemplos de Lenguajes star-free

- Σ^* es star-free.
  $r = \overline{\emptyset}$.
- $\{w \in \Sigma^* : w$ empieza con b o termina con $b\}$.
  $r = b\overline{\emptyset} + \overline{\emptyset}b$.
- $\{w \in \Sigma^* : w$ no contiene la subpalabra $aa\}$.
  $r = \overline{\overline{\emptyset}aa\overline{\emptyset}}$.

Semigrupos y Monoides

Un par $\langle S, \cdot \rangle$ es un semigrupo si $\cdot$ es una operación asociativa: $(x \cdot y) \cdot z = x \cdot (y \cdot z)$ para todo $x, y, z \in S$.

Un semigrupo $\langle M, \cdot \rangle$ es un monoide si tiene un elemento neutro e: $xe = ex = x$ para todo $x \in M$.

- Si el neutro e existe, es único.

Un monoide $\langle G, \cdot \rangle$ es un grupo si cada elemento $x \in G$ tiene un inverso $x^{-1} \in G$: $xx^{-1} = x^{-1}x = e$, donde e es el neutro.

Dado un semigrupo $\langle S, \cdot \rangle$ y $X \subseteq S$, decimos que S es un sub{semigrupo,monoide,grupo} si $\cdot$ es cerrado en X y X es un {semigrupo,monoide,grupo}.

Ejemplos de Semigrupos y Monoides

$\langle \mathbb{N} \setminus \{0\}, + \rangle$ es un semigrupo ($+$ es asociativa).

- $\langle \mathbb{N}, + \rangle$ es un monoide, pues 0 es neutro.
- $\langle \mathbb{Z}, + \rangle$ es un grupo, pues cada elemento tiene inverso.

$\mathbb{Z}_2 = \{0, 1\}$ es un monoide con la multiplicación módulo 2.

- 1 es neutro.

$\langle \mathbb{Z}, \cdot \rangle$ es un monoide (no hay inversos en general),
pero $\langle \{1\}, \cdot \rangle$ es un subgrupo.

Semigrupos Aperiódicos

Definición

Un semigrupo $\langle S, \cdot \rangle$ se dice aperiódico si
para cada $s \in S$ existe $n \in \mathbb{N}$ tal que $s^{n+1} = s^n$.

Teorema

Un semigrupo finito es aperiódico si y sólo si
no contiene subgrupos no triviales.

Ejemplos de Semigrupos Aperiódicos

El monoide $\langle \mathbb{Z}_2, \cdot \rangle$ es aperiódico. En efecto, $0^2 = 0^1$, y $1^2 = 1^1$.

$\langle \mathbb{Z}_2, + \rangle$ no es aperiódico, pues $1^{k+1} \neq 1^k$ para todo k.

Sea $M = \mathbb{N}$ el monoide con la operación $x \cdot y = \text{máx}\{x, y\}$.
M es aperiódico. En efecto, $x \cdot x = x$ para todo $x \in M$.

Ideal de un Semigrupo

Dado un semigrupo $\langle S, \cdot \rangle$ y $X \subseteq S$, se definen los conjuntos:

- $XS = \{x \cdot s : x \in X, s \in S\}$ y
- $SX = \{s \cdot x : x \in X, s \in S\}$.

Un subconjunto I de un semigrupo S es un ideal si $IS \subseteq I$ y $SI \subseteq I$.

Ejemplos de Ideales

$\emptyset$ y S son ideales triviales de un semigrupo $\langle S, \cdot \rangle$.

Ejercicio

Sea $S = \mathbb{N}$ el semigrupo con la operación $x \cdot y = \text{mín}\{x, y\}$.
¿Cómo son los ideales de S?

Si $I = \{0, \ldots, k\}$ para algún k, entonces I es un ideal de S.

El Ideal Prohibitivo de x

Sea M un monoide y $x \in M$.
El ideal prohibitivo (*forbidding ideal*) de x se define como

$$\begin{aligned} F_x &= \{y \in M : pyq \neq x \text{ para todo } p, q \in M\} \\ &= \{y \in M : x \notin MyM\} \end{aligned}$$

Debemos demostrar que F_x es un ideal: si $y \in F_x$ y $m, p, q \in M$,

$$p(ym)q = (p)y(mq) \neq x,$$

pues $y \in F_x$. Como p, q son arbitrarios, $x \notin M(ym)M$,
por lo que $ym \in F_x$. Como y, m son arbitrarios, $F_xM \subseteq F_x$.
Similarmente, $MF_x \subseteq F_x$, y se concluye que F_x es un ideal.

Ejemplo de Ideal Prohibitivo

Proposición

Sea $\langle S, \cdot \rangle$ un semigrupo aperiódico. Para todo $x, p, q \in S$, si $x = pxq$, entonces $x = px = xq$.

Si $\langle M, \cdot \rangle$ es un monoide aperiódico con neutro e, entonces $F_e = M \setminus \{e\}$.

En el monoide aperiódico $\langle \mathbb{Z}_2, \cdot \rangle$, $F_0 = \emptyset$ y $F_1 = \{0\}$.

La Relación $\equiv_L$

Dado un lenguaje L, se define la relación $\equiv_L$ sobre Σ^*:

$$x \equiv_L y \text{ si para todo } u \in \Sigma^*,\ xu \in L \Leftrightarrow yu \in L.$$

$\equiv_L$ es una relación de equivalencia.

Teorema (Nerode)

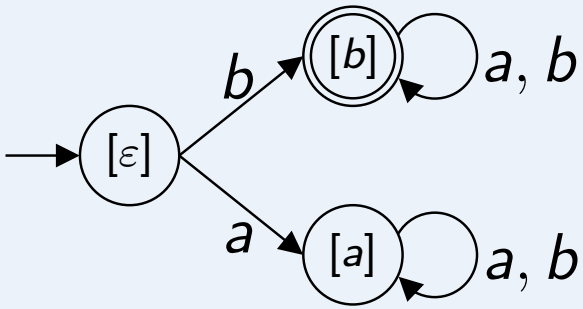
L es regular si y sólo si el índice de $\equiv_L$ es finito.

El Autómata Mínimo

El AFD mínimo se puede construir con $\Sigma^* / \equiv_L$.

Ejemplo

Sea L el lenguaje sobre $\{a, b\}$ de las palabras que empiezan con b. ¿Cuáles son las clases de equivalencia de $\equiv_L$?



Caracterización de $\equiv_L$ con Autómata Mínimo

Teorema

Sea $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ un autómata mínimo que acepta L. Entonces $x \equiv_L y$ si y sólo si $\hat{\delta}(q_0, x) = \hat{\delta}(q_0, y)$.

La Relación $\sim_L$

Dado un lenguaje L, se define la relación $\sim_L$ sobre Σ^*:

$$x \sim_L y \text{ si para todo } u, v \in \Sigma^*,\ uxv \in L \Leftrightarrow uyv \in L.$$

$\sim_L$ es una relación de equivalencia.

Teorema (Myhill)

L es regular si y sólo si el índice de $\sim_L$ es finito.

- En efecto, si $|\Sigma^*/\equiv_L| = n$, entonces $|\Sigma^*/\sim_L| \leq n^n$.

Caracterización de $\sim_L$ con Autómata Mínimo

Teorema

Sea $\mathcal{A} = (Q, \Sigma, \delta, q_0, F)$ un autómata mínimo que acepta L.
Entonces $x \sim_L y$ si y sólo si para todo $q \in Q$, $\hat{\delta}(q, x) = \hat{\delta}(q, y)$.

El Monoide Sintáctico de un Lenguaje

El conjunto $\Sigma^*/\sim_L$, equipado con la operación $[x][y] = [xy]$, constituye el monoide sintáctico de L, $\mathbf{Syn}(L)$.

La operación está bien definida (¿por qué?) y es asociativa:

$$([x][y])[z] = [xy][z] = [(xy)z] = [x(yz)] = [x][yz] = [x]([y][z])$$

Ejemplos de Monoide Sintáctico

Sea $\Sigma = \{a\}$ y $P = \{x \in \Sigma^* : |x| \text{ es par}\}$. ¿Cómo es $\mathbf{Syn}(P)$?

$$\mathbf{Syn}(P) = \left\{[\varepsilon]_{\sim_P}, [a]_{\sim_P}\right\}$$

La operación $\cdot$ está definida por

$\cdot$	$[\varepsilon]$	$[a]$
$[\varepsilon]$	$[\varepsilon]$	$[a]$
$[a]$	$[a]$	$[\varepsilon]$

Notamos que $\mathbf{Syn}(P)$ no es aperiódico: $[a]^{k+1} \neq [a]^k$ para todo k.

Ejemplos de Monoide Sintáctico

Sea $\Sigma = \{(,)\}$ y B el lenguaje de los paréntesis balanceados. Notemos que B no es regular. ¿Cómo es $\mathbf{Syn}(B)$?

$$\mathbf{Syn}(B) = \bigcup_{a,b \in \mathbb{N}} \{[)^a(^b]_{\sim_B}\}$$

No es aperiódico: $[))(]^{k+1} \neq [))(]^k$ para todo k.

Homomorfismos

Dados los monoides M_1, M_2, una función $h : M_1 \to M_2$ es un homomorfismo si $h(xy) = h(x)h(y)$ para todo $x, y \in M_1$.

Dado un monoide M y un alfabeto Σ, una función $h : \Sigma^* \to M$ es un homomorfismo si $h(xy) = h(x)h(y)$ para todo $x, y \in \Sigma^*$.

El Lenguaje Reconocido por un Monoide

Definición

Dado un alfabeto Σ y un monoide M, decimos que $L \subseteq \Sigma^*$ es reconocido por M si existe $X \subseteq M$ y un homomorfismo $h : \Sigma^* \to M$ tal que $L = h^{-1}(X)$.

Abusando notación, decimos que $L \subseteq \Sigma^*$ es reconocido por un homomorfismo $h : \Sigma^* \to M$ si existe $X \subseteq M$ tal que $h^{-1}(X)$ para algún $X \subseteq M$.

Paridad es Reconocido por $\langle \mathbb{Z}_2, + \rangle$

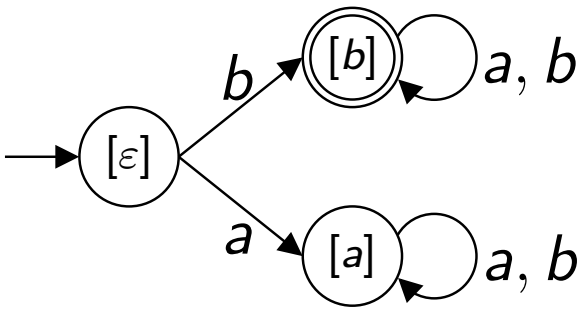
Consideremos el lenguaje $P = \{x \in \{a\}^* : |x| \text{ es par}\}$ y el monoide $M = \langle \{0, 1\}, + \rangle$, donde la adición es módulo 2.

Afirmación. P es reconocido por M.

En efecto, $h : \{a\}^* \to M$ dado por $h(x) = |x| \mod 2$ es un homomorfismo y $P = h^{-1}(\{0\})$.

Otro Lenguaje Reconocido por un Monoide

Sea $\Sigma = \{a, b\}$ y $L = L\left(b\overline{\emptyset}\right)$, es decir, el lenguaje sobre Σ de palabras que empiezan con b.



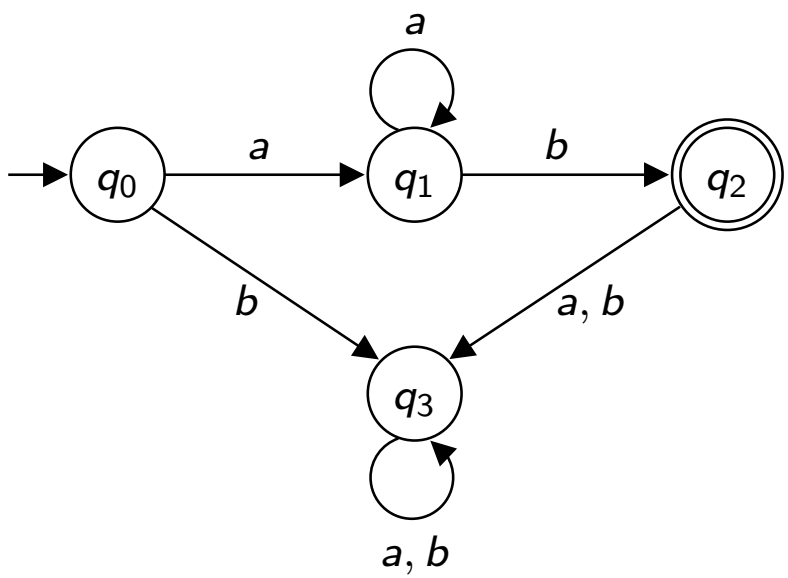
$\cdot$	$[\varepsilon]$	$[a]$	$[b]$
$[\varepsilon]$	$[\varepsilon]$	$[a]$	$[b]$
$[a]$	$[a]$	$[a]$	$[a]$
$[b]$	$[b]$	$[b]$	$[b]$

Notar que L es star-free, y que $\mathbf{Syn}(L)$ es aperiódico:

$$[\varepsilon]^2 = [\varepsilon], [a]^2 = [a], [b]^2 = [b].$$

Un Lenguaje más Reconocido por un Monoide

Sea $\Sigma = \{a, b\}$ y $L = L\left(aa^*b\right)$, es decir, el lenguaje sobre Σ de palabras que empiezan con una o más a y terminan con b.



Un Lenguaje más Reconocido por un Monoide

El monoide sintáctico de $L = L(aa^*b)$ tiene la siguiente operación:

$\cdot$	$[\varepsilon]$	$[a]$	$[b]$	$[ab]$	$[bb]$
$[\varepsilon]$	$[\varepsilon]$	$[a]$	$[b]$	$[ab]$	$[bb]$
$[a]$	$[a]$	$[a]$	$[ab]$	$[ab]$	$[bb]$
$[b]$	$[b]$	$[bb]$	$[bb]$	$[bb]$	$[bb]$
$[ab]$	$[ab]$	$[bb]$	$[bb]$	$[bb]$	$[bb]$
$[bb]$	$[bb]$	$[bb]$	$[bb]$	$[bb]$	$[bb]$

Notar que L es star-free: $L = L\left(\overline{b\overline{\emptyset} + \overline{\emptyset}a}\right)$. $\mathbf{Syn}(L)$ es aperiódico:

$$[\varepsilon]^2 = [\varepsilon], [a]^2 = [a], [b]^3 = [b]^2, [ab]^3 = [ab]^2, [bb]^2 = [bb].$$

El Teorema de Schützenberger

En 1965, Marcel Paul Schützenberger dio un importante paso en la Teoría Algebraica de Autómatas.

Teorema

L es star-free si y sólo si $\mathbf{Syn}(L)$ es finito y aperiódico.

El Teorema de Schützenberger

La demostración dada por Schützenberger es bastante técnica.
A continuación discutiremos las ideas involucradas.

Idea 1

Demostrar que **Syn**(L) es el monoide más pequeño que reconoce L.

Idea 2

Si $h : \Sigma^* \to M$ es un homomorfismo, $X \subseteq M$, entonces

$$h^{-1}(X) = \bigcup_{x \in X} h^{-1}(x).$$

El Teorema de Schützenberger

Idea 3

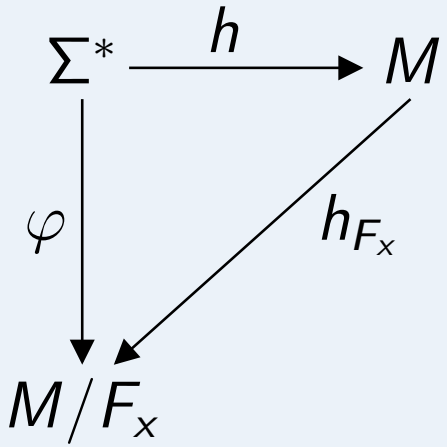
Como los lenguajes star-free son cerrados bajo unión, podemos buscar una expresión regular star-free para $h^{-1}(x)$, con $x \in X$ arbitrario, y luego operar estas expresiones con $+$.

Idea 4

Un ideal I de un monoide finito M determina un monoide M/I cuyo tamaño es igual a $|M| - |I| + 1 \leq |M|$, via un homomorfismo natural que denotaremos h_I.

Idea 5

Si $h : \Sigma^* \to M$ es un homomorfismo y $x \in M$, entonces $h^{-1}(x) = (h_{F_x} \circ h)^{-1}(x)$, donde F_x es el ideal prohibitivo de x.



Aquí, $\varphi = h_{F_x} \circ h$ es un homomorfismo, y $|M/F_x| = |M| - |F_x| + 1$. Así, si $|F_x| > 1$, el lenguaje $h^{-1}(x)$ puede ser reconocido por un monoide más pequeño que M.

Idea 6

Si $L = h^{-1}(I)$ para algún ideal I en un monoide finito y aperiódico M, entonces L puede ser generado por una expresión star-free que involucra lenguajes reconocidos por monoides aperiódicos más pequeños.

En el caso base, el ideal $I = \emptyset$ reconoce el lenguaje vacío, reconocido por la expresión $\emptyset$. En el paso inductivo se considera una palabra arbitraria $w \in L$ y se construyen expresiones de la forma

$$\overline{\emptyset} a h^{-1}(y) b \overline{\emptyset} \subseteq L,$$

donde $w = avb$ y $h(v) = y$. La inducción termina porque hay finitas elecciones para el triple (a, y, b).

El Teorema de Schützenberger

Idea 7

Sea M un monoide aperiódico. Entonces, para cada $x \in M$,

$$\{x\} = (xM \cap Mx) \setminus F_x.$$

Como consecuencia, podemos separar $h^{-1}(x)$:

$$h^{-1}(x) = \left(h^{-1}(xM) \cap h^{-1}(Mx)\right) \setminus h^{-1}(F_x)$$

El Teorema de Schützenberger

Idea 8

Sea M un monoide finito y aperiódico con neutro e, Σ un alfabeto finito y $h : \Sigma^* \to M$ un homomorfismo. Si $e \neq x \in M$, entonces existe $Y \subseteq M$ tal que, para todo $y \in Y$, $F_x \subsetneq F_y$ y $h^{-1}(x)$ puede ser expresado con una expresión star-free que involucra los lenguajes $h^{-1}(y)$ para $y \in Y$ y otros lenguajes reconocibles por monoides aperiódicos más pequeños.

Esbozo de la demostración

Por inducción en el tamaño del monoide. Con el monoide vacío podemos reconocer $\emptyset$, y con un monoide de un elemento podemos reconocer $\Sigma^* = \overline{\emptyset}$.

Para el paso inductivo basta con considerar $h^{-1}(x)$ para algún $x \in M$. Se expresa $h^{-1}(x)$ con una expresión que involucra $h^{-1}(y)$ para ciertos $y \in Y$ que satisfacen $F_x \subsetneq F_y$, de modo que $|M/F_y| < |M/F_x| \leq M$.

Como estos lenguajes $h^{-1}(y)$ son reconocidos por monoides más pequeños que M, podemos aplicar la hipótesis de inducción.