Langages Formels

TD 4

Isa Vialard
vialard@lsv.fr

March 4, 2024

Exercise 1 : Brzozowski-McCluskey algorithm

The goal of this exercise is to translate a finite automaton into a rational expression, giving an alternate proof of the associated implication of Kleene's theorem. We will proceed by successive transformations of the automaton.

1. We call *strongly normalized* every automaton which has a unique initial state to which no transition leads and a unique final state with no exiting transition, i.e. an automaton $\mathcal{A} = \langle Q, \Sigma, \{i\}, \{f\}, \delta\rangle$ such that for every state q and letter a, $(q, a, i) \notin \delta$ and $(f, a, q) \notin \delta$. Show that for all finite automaton, there is a strongly normalized automaton which recognizes the same language.

   $Q \leftarrow Q \cup \{i, f\}$, $I \leftarrow \{i\}$, $F \leftarrow \{f\}$,
   $\delta \leftarrow \delta \cup \{(i, a, q) | (q_i, a, q) \in \delta, q_i \in I\} \cup \{(q, a, f) | (q, a, q_f) \in \delta, q_f \in F\}$. Si il y a un état commun à I et F, on rajoute une ε-transition de i à f.

We will use a generalization of the definition of finite automata: the transition function will be a subset of $Q \times 2^{\Sigma^*} \times Q$. An execution of such an automaton recognizes the concatenation of languages of the transitions' labels. The automaton recognizes the union of the languages of all its accepting executions.

2. Show that every generalized automaton is equivalent to a generalized automaton in which there exists exactly one transition between each pair of states: $q' \in \delta(q, L)$ et $q' \in \delta(q, L')$ implies $L = L'$.

   On peut remplacer les transitions $q \xrightarrow{L} q'$ et $q \xrightarrow{L'} q'$ par une unique transition $q \xrightarrow{L+L'} q'$. S'il n'y a pas de transition $q \xrightarrow{L} q'$, on ajoute $q \xrightarrow{\emptyset} q'$.

3. Let $\mathcal{A}$ be a strongly normalized generalized automaton with initial state i and final state f. Let $q \in Q_{\mathcal{A}}$, $q \notin \{i, f\}$. Show that there exists an automaton equivalent to $\mathcal{A}$ with set of states $Q_{\mathcal{A}} \setminus \{q\}$.

   Pour tout $q_1, q_2 \neq q$, on peut remplacer les transitions $q_1 \xrightarrow{L_1} q$, $q \xrightarrow{L_3} q_2$, $q \xrightarrow{L_2} q$ et $q_1 \xrightarrow{K} q_2$ par une transition $q_1 \xrightarrow{L_1 \cdot L_2^* \cdot L_3 + K} q_2$.

4. Conclude that if L is recognized by a strongly normalized generalized automaton $\mathcal{A}$, then L belongs to the rational closure of the labels of the transitions of $\mathcal{A}$.

Si L est reconnu par un snga $\mathcal{A}$, alors il est reconnu par un snga à deux états $i \xrightarrow{L} f$ construit en suivant les deux questions précédentes. Donc L appartient à la fermeture par somme et concaténation et étoile des étiquettes de $\mathcal{A}$.

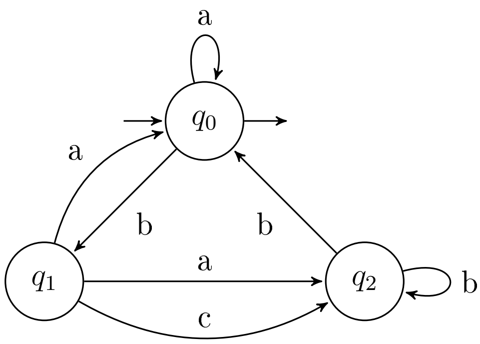
5. Show that every finite automaton has an equivalent generalized automaton.

On remplace chaque étiquette a par le langage $\{a\}$.

6. Give a procedure which, given a finite automaton, outputs a rational expression of same language.

On applique les questions 1,5,2,3.

7. Apply the construction to compute a rational expression corresponding to the following automaton:



$(a + b(a + (a + c)b^+)^*$

8. We consider the alphabet $\Sigma_n = [1; n] \times [1; n]$ and define:

$$L = \{ (a_1, a_2)(a_2, a_3) \dots (a_m, a_{m+1}) \ : \ m \geq 1, \ a_1 = 1, \ a_{m+1} = n \}$$

(a) Give an automaton of linear size recognizing L.

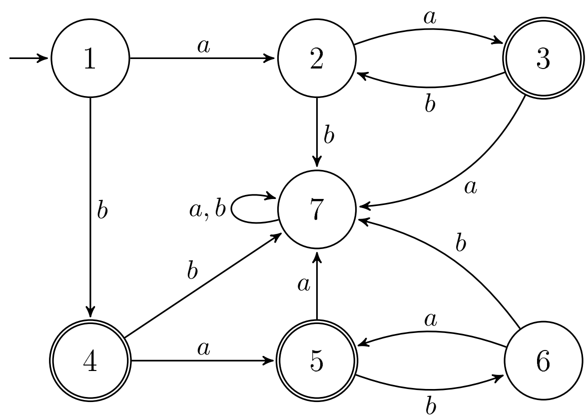
$Q = [1; n], I = \{1\}, F = \{n\}$ and $\delta = \{(a, (a, b), b) | a, b \in [1; n]\}$

(b) What is the size of the expression obtained by this construction on this automaton?

On rajoute 2 états i, f. Au début chaque transition est de taille 1 (ici la taille = nombres d'apparitions de lettres). On observe que quand on retire un état avec la question 3, la taille de chaque transition est multiplié par 4. Donc on a une expression de taille environ 4^n : c'est exponentiel.

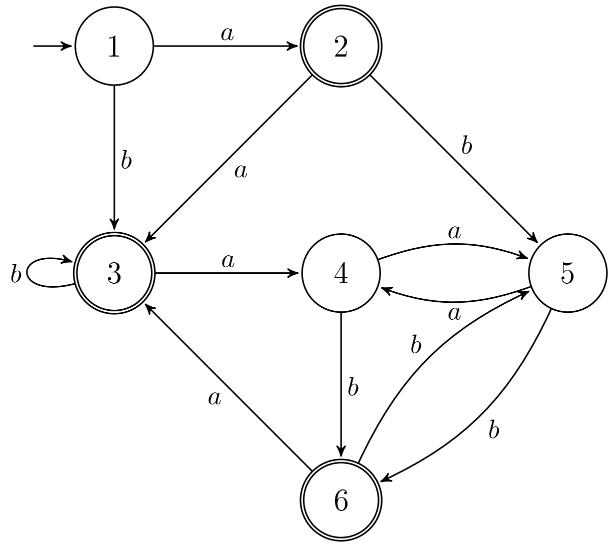
Exercise 2 : Minimization by Moore's algorithm

1. Minimize the automata $\mathcal{A}_1$ and $\mathcal{A}_2$, using Moore's algorithm:



(a) Automaton $\mathcal{A}_1$

$\mathcal{A}_1$: On peut unifier les états 2 et 6, et les états 3 et 5.
$\mathcal{A}_2$: On peut unifier les états 2 et 6, et les états 4 et 5.



(a) Automaton $\mathcal{A}_2$

2. Give a minimal automaton for $\mathcal{L} = ((a(a+b)^2 + b)^* a(a+b))^*$.

$\mathcal{L}$ a quatre résidus:

$$\begin{aligned}
\mathcal{L} &= ((a(a+b)^2 + b)^* a(a+b))^* = b^{-1}\mathcal{L} = (aab)^{-1}\mathcal{L} \\
a^{-1}\mathcal{L} &= (a+b)^2\mathcal{L} + (a+b)\mathcal{L} \\
(aa)^{-1}\mathcal{L} &= (a+b)\mathcal{L} + \mathcal{L} = (ab)^{-1}\mathcal{L} = (aaab)^{-1}\mathcal{L} \\
(aaa)^{-1}\mathcal{L} &= (a+b)^2\mathcal{L} + (a+b)\mathcal{L} + \mathcal{L} = (aaaa)^{-1}\mathcal{L}
\end{aligned}$$

Exercise 3 : Flashback

We proved that the language of palindromes over alphabet $\Sigma = \{a, b\}$ is not recognizable. We say a palindrome is *non trivial* if its length is greater than or equal to 2. Determine which of the following languages are recognizable (and prove it):

1. The language L_1 of words in Σ^* containing a non trivial palindrome as a prefix.

$L_1 = (ab^*a + ba^*b) \cdot \Sigma^*$ is recognizable.

2. The language L_2 of words in Σ^* containing a non trivial palindrome of even length as a prefix.

Not recognizable. Assume an automaton recognizes L_2, and consider the words $(ab)^n(ba)^n \in L_2$ for any $n \in \mathbb{N}$. By pigeonhole principle (principe des tiroirs), there is $m < n$ such that $\delta(i,(ab)^n) = \delta(j,(ab)^n)$. So $(ab)^n(ba)^m$ is accepted, contradiction.

Contrôle continu 4

À rendre pour le 29/02 à 16h15.

Exercise 4 : Résiduels

Calculer les résiduels de $\mathcal{L} = a^*(aa + b) + b(a + ba)^*$ et construire son automate minimal.

$\mathcal{L}$, $a^{-1}\mathcal{L} = a^*(a + b)$, $(aa)^{-1}\mathcal{L} = a^*b$ $(ab)^{-1}\mathcal{L} = \varepsilon$, $b^{-1}\mathcal{L} = (a + ba)^*$, $(bb)^{-1}\mathcal{L} = a(a + ba)^*$

Exercise 5 : Characterizing recognizability

We want to show a converse to the pumping lemma. We say that a language L satisfies P_h if for all $uv_1 \dots v_h w$ avec $|v_i| \geq 1$, there exists $0 \leq j < k \leq h$ such that

$$uv_1 \dots v_h w \in L \Leftrightarrow uv_1 \dots v_j v_{k+1} \dots v_h w \in L.$$

The theorem of Ehrenfeucht, Parikh & Rozenberg states that L is rational iff there exists h such that L satisfies P_h.

1. Show that if L satisfies P_h, then $w^{-1}L$ also does for every word $w \in \Sigma^*$.

Let $f = xv_1 \dots v_h y$. $f \in w^{-1}L$ iff $wf \in L$. If L satisfies P_h, then there exist $0 \leq i < j \leq h$ such that $wf \in L$ iff $wxv_1 \dots v_i v_{j+1} \dots v_h y \in L$, i.e. $xv_1 \dots v_i v_{j+1} \dots v_h y \in w^{-1}L$.

2. Let $h \in \mathbb{N}$. We want to show that the number of languages satisfying P_h is finite. We use the following statement of Ramsey's theorem:

For every k there is N such that, for every set E of cardinal greater than N and every bipartition $\mathcal{P}$ of $\mathfrak{P}_2(E) = \{\ \{e, e'\}\ :\ e, e' \in E,\ e \neq e'\ \}$, there exists a subset $F \subseteq E$ of cardinal k such that $\mathfrak{P}_2(F)$ is contained in one of the classes of $\mathcal{P}$.

Let N be the natural number given by Ramsey's theorem for $k = h + 1$. Let L and L' be two languages satisfying P_h and coinciding on words of size smaller than N. Prove that they coincide on words or size $M \geq N$, by induction on M. You may consider, for a word $f = a_1 \dots a_N t$ of size M (with $a_i \in \Sigma$), the following partition of $\mathfrak{P}_2([0; N])$:

$$X_f = \{\ (j, k)\ :\ 0 \leq j < k \leq N,\ a_1 \dots a_j a_{k+1} \dots a_N t \in L\ \}$$

$$Y_f = \mathfrak{P}_2([0; N]) \setminus X_f$$

Conclude.

According to Ramsay's theorem, for $E = [0; N]$, there exists $F = k_0 < \cdots < k_h \in E$ such that $\mathfrak{P}_2(F)$ is contained in either X_f or Y_f. Observe that we can express f as $uv_1 \ldots v_h w$ such that $u = a_0 \ldots a_{k_0-1}$, $v_i = a_{k_{i-1}} \ldots a_{k_i-1}$ and $w = a_{k_h} \ldots a_N t$. For all $0 \leq i < j \leq N$, let $f_{i,j} = uv_1 \ldots v_i v_{j+1} \ldots v_h w$.

If $\mathfrak{P}_2(F) \subseteq X_f$ then $f_{i,j} \in L$ for all i, j. If $\mathfrak{P}_2(F) \subseteq Y_f$ then $f_{i,j} \notin L$ for all i, j. In both cases, $f_{i,j} \in L$ iff $f_{i',j'} \in L$ for all pairs (i, j) and (i', j'). Since L and L' satisfy P_h then there exist i, j such that $f \in L$ iff$f_{i,j} \in L$, and there exist i', j' such that $f \in L'$ iff$f_{i',j'} \in L'$. Thus by induction hypothesis, $f \in L$ iff $f \in L'$.

3. Conclude that if a language L satisfies P_h for some h, then L is regular.

Un langage vérifiant P_h a un nombre fini de quotients à gauche, donc il est régulier. Correction provenant de `https://www.irif.fr/~carton/Enseignement/Complexite/ENS/Cours/pumping.html`