TD2 31 janvier 2018

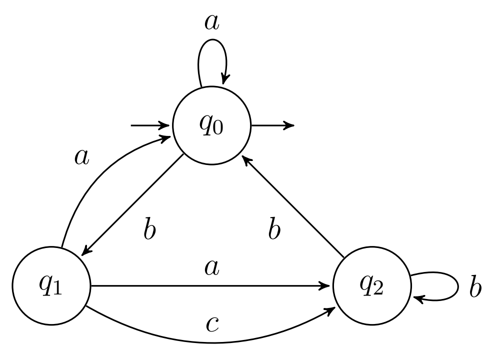
Langages Formels

Exercice 1 (Construction d'automates, ou pas)
Pour chacun des langages suivants sur $\Sigma = \{0, 1\}$, dire s'il est régulier (reconnaissable par un automate fini) ou pas. Justifier en donnant un automate ou un argument.

1. $\{ u \ : \ |u|_0 = |u|_1 \}$ où $|u|_a$ est le nombre d'occurrences de a dans u ;
2. $\{ 0^p 1^q \ : \ p \geq q \}$;
3. $\{ 0^p 1^q \ : \ p \geq q, \ q \leq 1337 \}$;
4. $\{ u \ : \ \overline{u}^2 = 0 \mod 3\}$ où $\overline{u}^2$ est l'interprétation de u en binaire, avec le bit de poids fort à droite.

Exercice 2 (Algorithme de Brzozowski-McCluskey)
L'objectif de cet exercice est de traduire un automate fini en une expression rationnelle. Nous allons procéder par transformations successives de l'automate, en utilisant une généralisation du type d'automate considéré : la relation de transition sera un sous-ensemble de $Q \times 2^{\Sigma^*} \times Q$. Une exécution d'un tel automate reconnait la concaténation des langages des transitions, et le langage reconnu par l'automate est l'union de ces exécutions.

1. Montrer que tout automate généralisé est équivalent à un automate généralisé pour lequel il existe exactement une transition entre chaque paire d'états : $q' \in \delta(q, L)$ et $q' \in \delta(q, L')$ implique $L = L'$.
2. Soit un automate généralisé $\mathcal{A}$ avec un unique état initial i, un unique état final f. Soit $q \in Q_{\mathcal{A}}$, $q \notin \{i, f\}$. Montrer qu'il existe un automate équivalent à $\mathcal{A}$ avec pour ensemble d'états $Q_{\mathcal{A}} \setminus \{q\}$.
3. En déduire que si L est reconnu par un automate généralisé $\mathcal{A}$, alors L appartient à la clôture rationnelle des étiquettes des transitions de $\mathcal{A}$.
4. Appliquer la construction pour calculer une expression rationnelle correspondant à l'automate suivant :



5. On considère l'alphabet $\Sigma_n = \{1, \ldots, n\} \times \{1, \ldots, n\}$, et on définit :

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

   (a) Donner un automate de taille quadratique reconnaissant L.
   (b) Quelle est la taille de l'expression obtenue par la construction précédente sur cet automate ?

Exercice 3 (Dérivées partielles, automate d'Antimirov)
La dérivée partielle est aux expressions rationnelles ce que le résiduel est aux langages : on va définir une opération $\partial_a(E)$ qui va correspondre (modulo interprétation) à $a^{-1}\mathcal{L}(E)$.

Formellement, pour une expression E et une lettre a, on définit l'*ensemble d'expressions* $\partial_a(E)$ comme suit, où l'opération de concaténation est naturellement étendue sur les ensembles d'expressions :

$$\begin{array}{rcl}
\partial_a(\underline{\emptyset}) & = & \partial_a(\underline{\varepsilon})\ =\ \emptyset \\
\partial_a(\underline{b}) & = & \left\{ \begin{array}{ll} \emptyset & \text{si } a \neq b \\ \{\underline{\varepsilon}\} & \text{sinon} \end{array} \right. \\
\partial_a(E + E') & = & \partial_a(E) \cup \partial_a(E') \\
\partial_a(E^*) & = & \partial_a(E) \cdot \{E^*\} \\
\partial_a(E \cdot E') & = & \left\{ \begin{array}{ll} \partial_a(E) \cdot \{E'\} & \text{si } \varepsilon \notin E \\ (\partial_a(E) \cdot \{E'\}) \cup \partial_a(E') & \text{sinon} \end{array} \right.
\end{array}$$

On définit ensuite $\partial_w(E)$ pour un mot w quelconque en posant $\partial_\varepsilon(E) = \{E\}$ et $\partial_{wa}(E) = \partial_a(\partial_w(E))$, où $\partial_w(S) = \bigcup_{E \in S} \partial_w(E)$ quand S est un ensemble d'expressions.

Étant donné un ensemble d'expression régulières S, on note $\mathcal{L}(S) = \bigcup_{E \in S} \mathcal{L}(E)$.

1. Calculer les dérivées partielles de $(ab+b)^*ba$ par a et b.
2. On note $w^{-1}L = \{u \in \Sigma^* \mid wu \in L\}$ le quotient à gauche de $L \subseteq \Sigma^*$ par $w \in \Sigma^*$.

   Montrer que pour tous $L, L' \subseteq \Sigma^*$ et $a \in \Sigma$,

$$\begin{aligned} a^{-1}(L \cup L') &= (a^{-1}L) \cup (a^{-1}L') \\ a^{-1}L^* &= (a^{-1}L) \cdot L^* \\ a^{-1}(L \cdot L') &= \begin{cases} a^{-1}L \cdot L' & \text{si } \varepsilon \notin L \\ (a^{-1}L \cdot L') \cup (a^{-1}L') & \text{sinon} \end{cases} \end{aligned}$$

3. Montrer que $\mathcal{L}(\partial_w(E)) = w^{-1}\mathcal{L}(E)$.
4. Utiliser cette construction pour traduire une expression rationnelle en un automate acceptant le même langage, en ne s'inquiétant pas de la finitude de cet automate pour l'instant.

   Appliquer cette construction à l'expression $(ab+b)^*ba$.
5. On définit l'ensemble des suffixes non vides d'un mot :

$$\mathrm{Suf}(w) = \{\ v \in \Sigma^+\ :\ \exists u, w = uv\ \}$$

   Montrer les résultats suivants pour $w \in \Sigma^+$:

$$\begin{aligned} \partial_w(E+E') &= \partial_w(E) \cup \partial_w(E') \\ \partial_w(E \cdot E') &\subseteq (\partial_w(E) \cdot E') \cup \bigcup_{v \in \mathrm{Suf}(w)} \partial_v(E') \\ \partial_w(E^*) &\subseteq \bigcup_{v \in \mathrm{Suf}(w)} \partial_v(E) \cdot E^* \end{aligned}$$

6. Soit $\|E\|$ le nombre d'occurrences de lettres de Σ dans E. Montrer que l'ensemble des dérivées partielles différentes de E contient au plus $\|E\|+1$ éléments.

Exercice 4 (Critère de reconnaissabilité)
On veut montrer que la version "ssi" de la troisième formulation du lemme de l'étoile est une caractérisation des langages réguliers. On dit qu'un langage

L satisfait P_N si pour tout $uv_1 \dots v_N w$ avec $|v_i| \geq 1$ il existe $0 \leq j < k \leq N$ tel que, pour tout $n \in \mathbb{N}$,

$$uv_1 \dots v_N w \in L \Leftrightarrow uv_1 \dots v_j (v_{j+1} \dots v_k)^n v_{k+1} \dots v_N w \in L.$$

Le théorème de Ehrenfeucht, Parikh & Rozenberg affirme que L est régulier ssi il existe un entier N tel que L satisfait P_N.

1. Montrer que si L appartient à P_N alors $a^{-1}L$ aussi, pour tout $a \in \Sigma$.
2. Démontrer le théorème, en supposant qu'il n'y a qu'un nombre fini de langages satisfaisant P_N pour un N fixé.
3. On rappelle le théorème de Ramsey, spécialisé pour nos besoins : pour tout N il existe R tel que, pour tout ensemble E de cardinal au moins R et pour toute partition $\mathcal{P}$ de $\mathfrak{P}_2(E) = \{ \{e, e'\} \ : \ e, e' \in E, \ e \neq e' \}$ en deux classes, il existe un sous-ensemble $F \subseteq E$ de cardinal N tel que $\mathfrak{P}_2(F)$ est tout entier contenu dans une seule classe de $\mathcal{P}$.

   Soient L et L' deux langages satisfaisant P_N et coïncidant sur les mots de longueur inférieure $R-1$. Montrer qu'ils coïncident aussi sur les mots de longueur $M \geq R-1$, par induction sur M. On pourra considérer, pour un $f = a_1 \dots a_{R-1} t$ de longueur M (avec $a_i \in \Sigma$) la partition suivante de $\mathfrak{P}_2([0; N-1])$:

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

   $$Y_f = \mathfrak{P}_2([0; R-1]) \setminus X_f$$

4. Conclure.