TD2 28 janvier 2016

Langages Formels

David Baelde
<baelde@lsv.ens-cachan.fr>

Exercice 1 (Des étoiles)
On considère les énoncés suivants, de plus en plus forts. Soit $L \in \mathrm{Rec}(\Sigma^*)$, il existe $N \geq 0$ tel que pour tout $x \in L$:

(a) Si $|x| \geq N$ alors il peut s'écrire $x = u_1u_2u_3$ avec $u_2 \neq \epsilon$ et $u_1u_2^*u_3 \subseteq L$.

(b) Si $x = w_1w_2w_3$ avec $|w_2| \geq N$ alors $w_2 = u_1u_2u_3$ avec $u_2 \neq \epsilon$ et $w_1u_1u_2^*u_3w_3 \subseteq L$.

(c) Si $x = uv_1v_2 \dots v_Nw$ avec $|v_i| \geq 1$ alors il existe $0 \leq j < k \leq N$ tel que $uv_1 \dots v_j(v_{j+1} \dots v_k)^*v_{k+1} \dots v_Nw \subseteq L$.

On sait que les langages réguliers satisfont ces trois propriétés.

1. Montrer que $\{\ u\ :\ u = \widetilde{u}\ \}$, c'est à dire le langage des palindromes, n'est pas régulier.
2. De même pour $\{\ (ab)^n(cd)^n\ \} \cup (\Sigma^*\{\ aa, bb, cc, dd, ac\ \}\Sigma^*)$.
3. Montrer que le langage suivant satisfait (c) mais n'est pas régulier : $\{\ udv\ :\ u, v \in \{a,b,c\}^*,\ u \neq v$ ou bien l'un des deux contient un carré $\}$.

Exercice 2 (Fermeture par morphisme)
Soient A et B deux alphabets et $f : A^* \to B^*$ un morphisme. Pour $L_A \subseteq A^*$ et $L_B \subseteq B^*$ on note :

$$f(L_A) = \{\ f(u)\ :\ u \in L_A\ \} \qquad f^{-1}(L_B) = \{\ u \in A^*\ :\ f(u) \in L_B\ \}$$

Dans les deux questions suivantes, on procédera par construction d'automate (déterministe ou pas) en partant d'un automate déterministe reconnaissant le langage considéré. On tâchera d'être efficace concernant la taille de l'automate produit.

1. Montrer que si L_A est reconnaissable alors $f(L_A)$ aussi.

2. Montrer que si L_B est reconnaissable alors $f^{-1}(L_B)$ aussi.

Exercice 3 (Expressions rationnelles)
En vrac :

1. Donner une expression reconnaissant le langage des mots ne contenant pas le facteur *ba*. Donner une expression pour le langage des mots ne contenant ni le facteur *aaa* ni *bbb*.
2. Calculer à partir d'une expression rationnelle une expression dont le langage est celui des suffixes du langage original (les mots qui sont suffixes d'un mot du langage original).
3. De même, calculer l'ensemble des premières lettres possibles, et l'ensemble des dernières lettres possibles.
4. Donner un langage non reconnaissable dont le langage des préfixes et le langage des suffixes soit reconnaissable.

Exercice 4 (k-reconnaissables)
Un ensemble d'entiers est k-reconnaissable si l'ensemble de ses écritures en base k est reconnaissable.

1. Montrer que l'ensemble des entiers divisibles par 2 est 3-reconnaissable.
2. Montrer que l'ensemble des puissances de 3 est 3-reconnaissable mais pas 2-reconnaissable. On pourra utiliser le lemme de l'étoile, et considérer ce qu'il implique asymptotiquement sur les nombres.

Exercice 5 (Fermeture par morphisme des mots dans les langages)
Soit f un morphisme de $(A^*, .)$ dans $(2^{B^*}, .)$, tel que $f(a)$ est régulier pour tout $a \in A$.

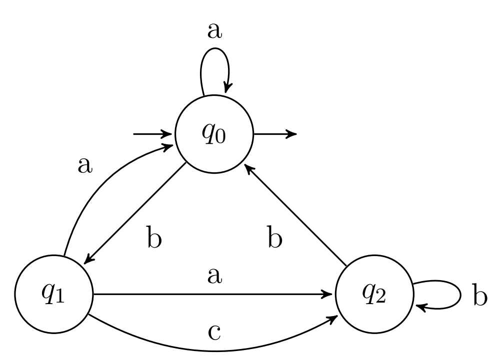
1. Soit $L \in \text{Rec}(A^*)$. Montrer que $f(L) = \cup_{u \in L} f(u)$ est régulier.
2. Soit un automate $\mathcal{A}$ sur A, et un automate $\mathcal{B}$ sur B. Montrer qu'on peut calculer, pour toute paire d'états s, t de $\mathcal{B}$ et pour tout état q de $\mathcal{A}$, l'ensemble suivant :

   $$R_{s,t}(q) = \{\ q' \ :\ \exists u.\ s \xrightarrow{u} t,\ q \xrightarrow{u} q'\ \}$$

   Évaluer la complexité de votre algorithme.
3. Soit $L \in \text{Rec}(B^*)$. Montrer que $f^{-1}(L) = \{\ u \in A^* \ :\ f(u) \subseteq L\ \}$ est régulier. On pourra commencer par montrer qu'on peut calculer, pour tout automate reconnaissant $f(a)$ et

Exercice 6 (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 fonction de transition sera de type $Q \times 2^{\Sigma^*} \to 2^Q$. 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 $\Sigma_n = [1; n] \times [1; n]$, et on définit :

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

   (a) 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 7 (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 $N-1$. Montrer qu'ils coïncident aussi sur les mots de longueur $M \geq N-1$, par induction sur M. On pourra considérer, pour un $f = a_1 \dots a_{N-1} t$ de longueur M (avec $a_i \in \Sigma$) la partition suivante de $\mathfrak{P}_2([0; N-1])$:

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

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

4. Conclure.