Langages formels

Devoir 1

à rendre le lundi 19 mars

Soit $\mathcal{F}$ un ensemble fini de symboles avec arités. Soit $\mathcal{X}$ un ensemble de variables. Un système de réécriture est un ensemble fini $\mathcal{S} \subseteq \mathsf{T}(\mathcal{F},\mathcal{X})^2$ de couples de termes. Étant donné un système de réécriture $\mathcal{S}$, on définit la relation de réduction $\rightarrow$ sur $\mathsf{T}(\mathcal{F})$ par $t \rightarrow u$ si et seulement s'il existe un contexte $c \in \mathsf{T}_\square(\mathcal{F})$, une substitution $\sigma : \mathcal{X} \rightarrow \mathsf{T}(\mathcal{F})$ et un couple $(g, d) \in \mathcal{S}$ tels que $t = c \cdot \sigma(g)$ et $u = c \cdot \sigma(d)$. On note $\rightarrow^*$ la clôture réflexive transitive de $\rightarrow$. Un terme $t \in \mathsf{T}(\mathcal{F})$ est dit irréductible s'il n'existe aucun u tel que $t \rightarrow u$. On note $\mathrm{Irr}(\mathcal{S})$ l'ensemble des termes (clos) irréductibles pour $\mathcal{S}$. Étant donné un langage d'arbres $L \subseteq \mathsf{T}(\mathcal{F})$, on définit

$$\begin{aligned} \mathcal{S}^*(L) &:= \{\, t \mid t \in \mathsf{T}(\mathcal{F}),\ \exists u \in L.\ u \rightarrow^* t \,\} && \text{réduits de } L \\ \mathcal{S}(L) &:= \mathcal{S}^*(L) \cap \mathrm{Irr}(\mathcal{S}) && \text{formes } \mathcal{S}\text{-normales de } L \end{aligned}$$

1. Montrer que si $\mathcal{S}$ est linéaire à gauche, c'est-à-dire si pour tout $(g, d) \in \mathcal{S}$, g est linéaire, alors $\mathrm{Irr}(\mathcal{S})$ est reconnaissable.

Correction On commence par prouver que pour tout terme linéaire $t \in \mathsf{T}(\mathcal{F},\mathcal{X})$, l'ensemble $R(t)$ des termes ayant une instance de t comme sous-terme est reconnaissable. Pour cela, on pose $Q = \mathrm{Pos}(t)$ l'ensemble des positions dans t. On définit la table de transition comme l'ensemble des transitions suivantes, pour tout $a \in \mathcal{F}_k$:

$$\begin{aligned} \langle q_1, \ldots, q_k\rangle, a &\rightarrow q && \text{si } t|_q \in \mathcal{X} \\ \langle q.1, \ldots, q.k\rangle, a &\rightarrow q && \text{si } t|_q = a(t|_{q.1}, \ldots, t|_{q.k}) \\ \langle q_1, \ldots, q_k\rangle, a &\rightarrow \varepsilon && \text{si } \varepsilon \in \{q_1, \ldots, q_k\} \end{aligned}$$

En prenant ε comme seul état final, on vérifie facilement que l'automate obtenu reconnaît $G(t)$. On a $\mathrm{Irr}(\mathcal{S}) = \mathsf{T}(\mathcal{F}) \setminus \bigcup_{(g,d)\in\mathcal{S}} G(g)$. Par hypothèse $\mathcal{S}$ est fini et linéraire à gauche, la reconnaissabilité est préservée par union finie et complémentaire, donc $\mathrm{Irr}(\mathcal{S})$ est reconnaissable.

2. Montrer que le nombre d'états d'un automate déterministe reconnaissant $\mathrm{Irr}(\mathcal{S})$ peut être parfois minoré par 2^{n-1}, n étant la somme des tailles des membres gauches de $\mathcal{S}$.

3. En général, si $\mathcal{S}$ n'est pas linéaire à gauche, $\mathrm{Irr}(\mathcal{S})$ est-il reconnaissable ? Supposons que l'on sache décider, étant donnés un langage reconnaissable $L \in \mathsf{T}(\mathcal{F})$ et un morphisme $h : \mathsf{T}(\mathcal{F}) \rightarrow \mathsf{T}(\mathcal{F})$, si $h(L)$ est reconnaissable. Montrer qu'on peut alors décider si $\mathrm{Irr}(\mathcal{S})$ est reconnaissable.

On se place dans le cas $\mathcal{F} = \{f(2), g(1), h(1), a(0)\}$.

4. On considère le langage $L = \{\, f(t,u) \mid t, u \in \mathsf{T}(\{g(1), h(1), a(0)\}) \,\}$, et le système de réécriture $\mathcal{S}$ constitué des règles suivantes :

$$\begin{aligned} f(g(x), h(y)) &\rightarrow f(x, y) & \qquad f(h(x), g(y)) &\rightarrow f(x, y) \\ g(h(x)) &\rightarrow x & h(g(x)) &\rightarrow x \\ f(a, x) &\rightarrow x & f(x, a) &\rightarrow x \end{aligned}$$

Les langages L, $\mathcal{S}^*(L)$ et $\mathcal{S}(L)$ sont-ils reconnaissables ?

Correction L est clairement reconnaissable. On remarque facilement que $\mathcal{S}^*(L)$ est exactement $L \cup \mathsf{T}(\{g(1), h(1), a(0))$, qui est lui aussi reconnaissable. Enfin, $\mathcal{S}$ est linéaire à gauche donc $\mathrm{Irr}(\mathcal{S})$ est reconnaissable et $\mathcal{S}(L)$ aussi. On peut en donner une description explicite : $\mathcal{S}(L) = g^*(a) \cup h^*(a) \cup f(g^+(a), g^+(a)) \cup f(h^+(a), h^+(a))$.

5. On considère le langage $L = \{ g(h^n(a)) \mid n \geqslant 0 \}$ et le système de réécriture $\mathcal{S}$ consititué de l'unique règle $g(x) \to f(x,x)$. Les langages L, $\mathcal{S}^*(L)$ et $\mathcal{S}(L)$ sont-ils reconnaissables ?

   Correction L est clairement reconnaissable. On a $S^*(L) = L \cup \{ f(h^n(a), h^n(a) \mid n \geqslant 0 \}$ qui n'est pas reconnaissable (lemme d'itération). On a $S(L) = \{ f(h^n(a), h^n(a) \mid n \geqslant 0 \}$ et ce langage n'est pas reconnaissable non plus.

On se place à nouveau dans le cas général. On appelle préservation de la reconnaissabilité la propriété selon laquelle, pour tout L reconnaissable, $\mathcal{S}^*(L)$ est reconnaissable.

6. On considère un système réécriture $\mathcal{S}$ qui préserve la reconnaissabilité. Que dire des deux problèmes de décision suivants ?
   – Étant donnés $u, v \in T(\mathcal{F})$, est-ce que $v \in \mathcal{S}^*(\{u\})$?
   – Étant donnés $L, L' \subseteq (\mathcal{F})$ reconnaissables, est-ce que $L' \subseteq \mathcal{S}^*(L)$?

7. On suppose que $\mathcal{S}$ est linéaire et monadique, c'est-à-dire que, pour toute règle $(g, d) \in \mathcal{S}$, g et d sont linéaires, g est de hauteur au moins 1 et d est soit une variable soit un terme $f(x_1, \dots, x_k)$ avec $f \in \mathcal{F}_k$ ($k \geqslant 0$) et les x_i tous distincts. Montrer que $\mathcal{S}$ préserve la reconnaissabilité.

   Correction On considère un automate $A = (Q, \Delta, F)$, non déterministe éventuellement avec ε-transitions, qui reconnaît L. Notons T_k l'ensemble des termes sur $\mathcal{F}$ à k trous, c'est-à-dire l'ensemble des termes linéaires de $T(\mathcal{F}, \{x_1, \dots, x_k\})$. On étend la relation de transition Δ en une relation $\Delta^* \subseteq Q^k \times T_k \times Q$ de façon naturelle. On étend la relation de réduction sur ces termes en supposant que les x_i n'apparaissent pas dans les règles du système. On définit l'automate $A' = (Q, \Delta', F)$, non déterministe avec ε-transitions, par

   $$\begin{aligned}\Delta' := &\{ \langle q_1, \dots, q_k \rangle, f, q \mid \exists t \in T_k, t \to^* f(x_1, \dots, x_k) \wedge (\langle q_1, \dots, q_k \rangle, t, q) \in \Delta^* \} \\ &\cup \{ q, \varepsilon, q' \mid \exists t \in T_k, \exists q_1, \dots, q_k \in Q, t \to^* x_i \wedge q_i = q \wedge (\langle q_1, \dots, q_k \rangle, t, q') \in \Delta^* \}\end{aligned}$$

   Reste à vérifier que cet automate reconnaît bien $S^*(L)$.

8. En déduire que si $\mathcal{S}$ est linéaire et monadique, alors pour tout L reconnaissable $\mathcal{S}(L)$ est reconnaissable.

   Correction Conséquence de la question précédente et du fait que $\mathrm{Irr}(\mathcal{S})$ est reconnaissable si $\mathcal{S}$ est linéaire à gauche.

9. Montrer que la préservation de la reconnaissabilité n'est plus garantie si l'on supprime la condition de linéarité à droite.