Des graphes de Myhill et des langages locaux

Benjamin Hellouin

Définition 1. *Un graphe de Myhill M est la donnée de (Σ, E, I, F), où :*
- *Σ est à la fois l'alphabet d'entrée et l'ensemble des états ;*
- *$E \subseteq \Sigma^2$ est l'ensemble des arêtes (et (Σ, E) est un graphe orienté) ;*
- *$I \subseteq \Sigma$ est l'ensemble des états initiaux ;*
- *$F \subseteq \Sigma$ est l'ensemble des états finaux.*

Le langage reconnu par un graphe de Myhill M, noté $\mathcal{L}(M)$, est l'ensemble des mots $u = u_1 \dots u_n \in \Sigma^* \backslash \epsilon$ tels que $u_1 \in I$, $u_n \in F$ et $\forall i, (u_i, u_{i+1}) \in E$.

1 Automates locaux standards

Définition 2. *Soit $A = (E, \Sigma, \delta, i, F))$ un automate fini déterministe. A est dit local si, pour tout $a \in \Sigma$, $\big|\{\delta(e,a) \mid e \in E\}\big| \leq 1$, et il est dit standard si $i \notin Im(\delta)$.*

Théorème 1. *Un langage L est reconnaissable par un graphe de Myhill si et seulement si il est reconnaissable par un automate fini local standard.*

Preuve. ($\Rightarrow$) Soit $M = (\Sigma, E, I, F)$ un graphe de Myhill. On construit l'automate A comme suit : l'alphabet est Σ, l'ensemble des états est $\Sigma \cup a_0$, l'état initial est $\{a_0\}$ et l'ensemble d'états finaux est F. La fonction de transition est définie comme suit :
- $\forall a \in I, \delta(a_0, a) = a$, non définie sinon.
- $\forall (a,b) \in \Sigma^2, \delta(a,b) = b$, non définie sinon.

A est déterministe, local et standard par construction, et il est facile de voir que $\mathcal{L}(M) = \mathcal{L}(A)$.

($\Leftarrow$) Soit $A = (Q, \Sigma, \delta, i, F)$ un automate déterministe local standard. On note q_a l'état tel que $\exists e, \delta(e,a) = q_a$, s'il existe (il est unique par localité). Alors on construit M comme suit : l'alphabet est Σ, et $E = \{(a,b) \mid \delta(q_a, b) = q_b\}, I = \{a \in \Sigma \mid \delta(i,a) \text{ est définie}\}$ et $F' = \{a \in \Sigma \mid q_a \in F\}$.

Supposons que $u = u_1 \dots u_n$ est reconnaissable par M. Alors $\delta(i, u_1)$ est définie et vaut q_{u_1}, et à chaque étape, $\delta(q_{u_i}, u_{i+1})$ existe et vaut $q_{u_{i+1}}$ par construction. Comme $q_{u_n} \in F$, u est reconnaissable par A. On peut faire le raisonnement inverse et montrer que $\mathcal{L}(M) = \mathcal{L}(A)$.

2 Lien avec les langages

Théorème 2. *Les langages reconnaissables par un graphe de Myhill sont exactement les langages locaux, i.e. les langages pouvant être décrits par une expression régulière de la forme $(P\Sigma^* \cap \Sigma^* S) \backslash \Sigma^* N \Sigma^*$ (où $P, S \subseteq \Sigma, N \subseteq \Sigma^2$).*

Vérifier l'appartenance d'un mot à un langage local revient à parcourir le mot en vérifiant ('localement') que les sous-mots de taille 2 n'appartiennent pas à un ensemble de sous-mots interdits.

Preuve. Il est évident que tout langage reconnaissable par un graphe de Myhill (Σ, E, I, F) s'écrit sous la forme $(I\Sigma^* \cap \Sigma^* F) \backslash \Sigma^* E \Sigma^*$. A l'inverse, ceci nous permet de définir un graphe de Myhill reconnaissant tout langage local.