Automates et Calculabilité

– Homomorphismes & Lemme de gonflement –

Leçon 7 : 1 dec 2003

Uwe Nestmann

EPFL-I&C

Mot-homomorphismes

Un mot-homomorphisme est une fonction totale $h : \Sigma^* \to \Gamma^*$ telle que pour tout $x, y \in \Sigma^*$

$$\begin{aligned} h(xy) &\equiv h(x)h(y) \\ h(\epsilon) &\equiv h(\epsilon) \end{aligned}$$

Pour définir un mot-homomorphisme, il suffit de déterminer son effet sur les symboles de l'alphabet Σ,
c'est-à-dire de définir $h : \Sigma \to \Gamma^*$.

Exemple:

Soit $\Sigma = \{\, a, b \,\}$ et $\Gamma = \{\, c, d, e \,\}$.
Soit $h(a) \stackrel{\text{def}}{=} ccc$ et $h(b) \stackrel{\text{def}}{=} dd$.
Alors $h(abaab) = \ldots$

Propriétés de clôture

Définition:
Soit $A \subseteq \Sigma^*$ et $B \subseteq \Gamma^*$.

$$h(A) \stackrel{\text{def}}{=} \{\, h(x) \mid x \in A \,\} \subseteq \Gamma^* \quad \text{est l'}\textbf{\textit{image}} \text{ de } A \text{ sous } h$$
$$h^{-1}(B) \stackrel{\text{def}}{=} \{\, x \mid h(x) \in B \,\} \subseteq \Sigma^* \quad \text{est la }\textbf{\textit{pré-image}} \text{ de } B \text{ sous } h$$

Théorème:
Soit $h : \Sigma^* \to \Gamma^*$ un homomorphisme.

1. Si $B \subseteq \Gamma^*$ est régulier, alors $h^{-1}(B)$ est aussi régulier.
2. Si $A \subseteq \Sigma^*$ est régulier, alors $h(A)$ est aussi régulier.

NB: Soit $A \subseteq \Sigma^*$.
Si $h(A)$ est régulier, alors A ne l'est pas forcément.

Exemple: Soit $\Sigma = \Gamma = \{\, a, b \,\}$, $A = \{\, a^n b^n \mid n \geq 0 \,\}$.
Comparer A avec $h^{-1}(h(A))$ pour et $h(a) \stackrel{\text{def}}{=} a$ et $h(b) \stackrel{\text{def}}{=} a$.

Exemple

Soit $N = (\mathcal{Q}, \Sigma \cup \{\boldsymbol{\epsilon}\}, \Delta, S, F)$ un AFN$_{\boldsymbol{\epsilon}}$.

$L(N)$ est le langage de N en utilisant la fonction $\widehat{\Delta}$ qui traite le symbole $\boldsymbol{\epsilon}$ comme s'il était un symbole standard.

$L_{\boldsymbol{\epsilon}}(N)$ est le langage de N en utilisant la fonction $\widehat{\Delta}_{\boldsymbol{\epsilon}}$ qui est basée sur la notion de $\boldsymbol{\epsilon}$-fermeture (introduite dans la leçon 5).

On peut définir $L_{\boldsymbol{\epsilon}}(N)$ en fonction de $L(N)$
en utilisant l'homomorphisme

$$\begin{array}{rcll} h : (\Sigma \cup \{\boldsymbol{\epsilon}\})^* & \to & \Sigma^* & \\ h(a) & \stackrel{\text{def}}{=} & a & \qquad \text{pour tout } a \in \Sigma \\ h(\boldsymbol{\epsilon}) & \stackrel{\text{def}}{=} & \epsilon & \end{array}$$

et donc, $L_{\boldsymbol{\epsilon}}(N) = h(L(N))$.

« Pigeonhole principle »

Aussi connu sous le nom de
« Fubini principle » :

« Étant donné en moyenne a enveloppes par trou,
il existe au moins un trou qui contient ***au moins*** a enveloppes,
et il existe au moins un trou qui contient ***au plus*** a enveloppes.
»

Une variante est connue sous le nom de
« Dirichlet's box principle » :

« Étant donné n boîtes et plus de n objets,
il existe au moins une boîte qui contient ***plus d'un*** objet. »

$A = \{ a^n b^n \mid n \geq 0 \}$ est-il régulier ?

Existe-t-il un automate fini (D,N ou N$_{\boldsymbol{\epsilon}}$) M
qui accepte ce langage A ?

Problèmes :

- □ Pout tout n, l'automate doit accepter $a^n b^n$.
- □ L'automate doit « se souvenir » du nombre n de a lorsqu'il passe le point central d'un mot et ce pour tout n.
- □ Le nombre d'états doit être fini.

Démontrer par l'absurde qu'il ne peut pas exister
un automate fini (D,N ou N$_{\boldsymbol{\epsilon}}$) M qui accepte ce langage A.

$A = \{ a^n b^n \mid n \geq 0 \}$ n'est pas régulier !

Soit $M = (\mathcal{Q}, \{a, b\}, \delta, s, F)$ un AFD tel que $L(M) = A$.

Nous utilisons le « pigeonhole principle »
sur le nombre d'états de M (dénoté par $k \in \mathbb{N}$), et
par rapport au paramètre n.

Choisissons $n \gg k$, par exemple $n = 2 * k + 42$.

Il existe au moins 1 état p dans lequel M doit passer plusieurs fois (en acceptant au moins un a entre les deux passages) dans la partie a^n pour pouvoir accepter mot $a^n b^n$.

Découpons le mot $a^n b^n$ en 3 parties déterminées par cet état p.
Soit u, v, w tel que $uvw = a^n b^n$ et

$$\widehat{\delta}(s, u) = p \qquad \widehat{\delta}(p, v) = p \qquad \widehat{\delta}(p, w) = r \in F.$$

Alors, M accepte tous les mots $u\underline{\underline{a}}^i vw$ pour $i \geq 0$!!
Donc, M n'accepte pas le langage A.
Donc, il ne peut pas exister un tel automate du tout.

$B = \{ a^{2^n} \mid n \geq 0 \}$ n'est pas régulier !

Nous démontrons par l'absurde qu'il ne peut pas exister un AFD $N = (\mathcal{Q}, \{a, b\}, \delta, s, F)$ qui accepte B :

- Prendre un mot a^{2^n} dans B qui est « suffisamment grand » par exemple pour $n \gg \|\mathcal{Q}\| = k$.
- Déduire par le « pigeonhole principle » l'existence d'un état p qui est traversé plusieurs fois durant la lecture des n premiers symboles a.
- Découper le mot a^{2^n} en 3 parties de longueurs i, j, k et satisfaisant $2^n = i + j + k$ et $0 < j \leq n$ tel que :

$$\widehat{\delta}(s, a^i) = p \qquad \widehat{\delta}(p, a^j) = p \qquad \widehat{\delta}(p, a^k) = r \in F.$$

- « Gonfler le boucle pour sortir du langage » :
M accepte tous les mots $a^i \underline{\underline{a^{j*m}}} a^k$ pour $m \geq 0$!!
Mais, par exemple pour $m = 1$, $a^{2^n + j} \notin B$.

Lemme de gonflement (Kozen)

Soit A régulier. Alors A satisfait la propriété $\boxed{P_K(A)}$:

Il existe $n \geq 0$ tel que
  pour tous les mots x, y, z tels que $xyz \in A$ et $|y| \geq n$,
  il existe des mots u, v, w tels que
    $y = uvw$, $v \neq \epsilon$,
    et pour tout $k \geq 0$, le mot $xuv^kwz \in A$.

La contraposée $\boxed{\neg P_K(A)}$ est utile :

Pour tout $n \geq 0$
  il existe des mots x, y, z tels que $xyz \in A$ avec $|y| \geq n$,
  et pour tous les mots u, v, w tels que
    $y = uvw$, $v \neq \epsilon$,
    il existe un $k \geq 0$ tel que le mot $xuv^kwz \notin A$.

Si $\neg P_K(A)$, alors A n'est pas régulier.

Lemme de gonflement (Hopcroft, Ullman)

Soit A régulier. Alors A satisfait la propriété $\boxed{P_{HU}(A)}$:

Il existe $n \geq 0$ tel que
  pour tous les mots $y \in A$ avec $|y| \geq n$,
  il existe des mots u, v, w tels que
    $y = uvw$, $v \neq \epsilon$, $|uv| \leq n$,
    et pour tout $k \geq 0$, le mot $uv^k w \in A$.

La contraposée $\boxed{\neg P_{HU}(A)}$ est utile :

Pour tout $n \geq 0$
  il existe un mot $y \in A$ tels que $|y| \geq n$,
  et pour tous les mots u, v, w tels que
    $y = uvw$, $v \neq \epsilon$, et $|uv| \leq n$,
    il existe un $k \geq 0$ tel que le mot $uv^k w \notin A$.

Si $\neg P_{HU}(A)$, alors A n'est pas régulier.

Si A est régulier, alors $P_{HU}(A)$.

Soit $M = (\mathcal{Q}, \Sigma, \delta, s, F)$ un AFD avec $L(M) = A$.
Soit $n = \|\mathcal{Q}\|$ le nombre d'états de M.
Prendre un mot arbitraire $y \in A$ tels que $|y| = m \geq n$.
Dénotons par $y = a_1 a_2 \cdots a_m$ les symboles de y.
Dénotons par $p_i = \widehat{\delta}(s, a_1 a_2 \cdots a_i)$ l'état atteint après le $i^{\text{ème}}$ symbole.

Par le « pigeonhole principle », comme nous n'avons que n états, alors il existe $0 \leq i < j \leq n$ tel que $p_i = p_j$.

$$u = a_1 \cdots a_i \qquad v = a_{i+1} \cdots a_j \qquad w = a_{j+1} \cdots a_m$$

Remarquez que $v \neq \epsilon$ parce que $i < j$.

Par la définition inductive de $\widehat{\delta}$, et pour $k \geq 0$, nous arrivons à :

$$\widehat{\delta}(s, u) = p_i \qquad \widehat{\delta}(s, uv) = p_j = p_i \qquad \widehat{\delta}(s, uvw) = p_m \in F$$

$$\widehat{\delta}(s, uv^k) = p_j = p_i \qquad \widehat{\delta}(s, uv^k w) = p_m \in F$$

CQFD

Non-regularité par « réduction » (I)

Nous pouvons aussi utiliser les propriétés de clôture
pour démontrer la non-regularité de certains langages.

Soit $f : 2^{\Sigma} \to 2^{\Sigma}$ une opération sur les langages sur Σ
qui préserve la propriété de régularité.
(C'est-à-dire : si A est régulier, alors $f(A)$ est aussi régulier.)

Supposons que nous nous intéressons à la question
de savoir si un langage B est régulier ou non.

Si nous avons déjà démontré que $f(B)$ est non-régulier,
et si f préserve la propriété de régularité,
alors B est aussi non-régulier.

Non-regularité par « reduction » (II)

Exemple:

Soit $D = \{ x \in \{ a, b \}^* \mid \#a(x) = \#b(x) \}$.
Supposons que D est régulier.

Alors, le langage $D \cap L(a^*b^*)$ doit aussi être régulier.

Cependant $D \cap L(a^*b^*) = \{ a^n b^n \mid n \geq 0 \}$,
dont nous avons démontré la non-régularité.

Nous avons une contradiction.
Donc, D ne peut pas être régulier.

Exemple: $\{ a^n b^m \mid n \geq m \}$.

Utiliser le fait que $A^I \overset{\text{def}}{=} \{ x^I \mid x \in A \}$ préserve la régularité.
(Nous avons rencontré l'opération $(-)^I$ dans la Série 3.)

Périodicité ultime (I)

Soit U un sous-ensemble des nombres naturels, alors $U \subseteq \mathbb{N}$.

U est *finalement périodique*
s'il existe des nombres $n \geq 0$ et $p \geq 0$
tel que pour tout $m \geq n$,
$m \in U$ si et seulement si $m + p \in U$.
Dans ce cas, p est appelé une période de U.

Exemple:

$$\{ 0, 3, 7, 11, 19, 20, 23, 26, 29, 32, 35, 38, 41, 44, 47, 50, \ldots \}$$

Vérifier le cas $n_1 = 20$ et $p_1 = 3$.
Vérifier le cas $n_2 = 21$ et $p_2 = 6$.
Vérifier le cas $n_3 = 100$ et $p_2 = 33$.

Périodicité ultime (II)

<u>**Théorème:**</u>
Soit $\Sigma = \{ a \}$ et $A \subseteq \Sigma^*$ de taille infinie.
A est régulier si et seulement si l'ensemble $\{ m \mid a^m \in A \}$ des longueurs des mots de A est finalement périodique.

Idée de la preuve : (« p-boucles avec n-queues »)

$\Rightarrow$ Soit $M = (\mathcal{Q}, \Sigma, \delta, s, F)$ un AFD avec $L(M) = A$.
Alors, pour tout $p \in \mathcal{Q}$, $|\{ \delta(p, a) \mid a \in \Sigma \}| = 1$.
Observer la forme du graphe des transitions de M.
Pour tout $m \geq n$, $a^m \in L(M)$ ssi $a^{m+p} \in L(M)$.

$\Leftarrow$ Soit U un ensemble finalement périodique avec n et p.
Construire l'automate $M = (\mathcal{Q}, \{a\}, \delta, \{0\}, \mathcal{Q} \cap U)$ avec $\mathcal{Q} = \{ 0, 1 \ldots, n+p-1 \}$ et $\delta(i) = n+((i-n)+_p 1)$.
Vérifier que $a^m \in L(M)$ ssi $a^{m+p} \in L(M)$ pour tout $m \geq n$.

Non-regularité par « reduction » (III)

<u>Corollaire:</u>
Soit A régulier sur un alphabet Σ fini.
Alors, **Len**$(A) \stackrel{\text{def}}{=} \{\, |x| \mid x \in A \,\}$ est finalement périodique.

Idée de la preuve :
Nous utilisons le mot-homomorphisme

$$\begin{aligned} h : \Sigma &\to \{a\} \\ b &\mapsto a \qquad\qquad \text{pour tout } b \in \Sigma \end{aligned}$$

pour arriver à $h(x) = a^{|x|}$.
Observer que **Len**$(h(A)) =$ **Len**(A).
Par la clôture de la régularité sous homomorphisme,
$h(A)$ est régulier sur l'alphabet $\{a\}$.
Alors **Len**$(h(A))$ est finalement périodique.