Option Informatique en Spé MP et MP*

Devoir à rendre après les vacances de Noël

Langages locaux, automates locaux
et algorithme de McNaughton, Yamada et Glushkov
(d'après Jean Berstel et Jean-Éric Pin)

Résumé

Le théorème de KLEENE affirme l'équivalence entre *reconnaissabilité* et *rationalité*. On connaît des preuves *constructives* de ce théorème : une telle preuve fournit un algorithme associant à un automate $\mathcal{A}$ une expression rationnelle e qui décrit le langage reconnu par $\mathcal{A}$, ou, en sens inverse, associant à une expression rationnelle e un automate $\mathcal{A}$ qui reconnaît le langage décrit par e.

C'est à cette deuxième construction, d'utilisation fréquente, que nous nous intéresserons. La méthode la plus connue, et aussi la plus facile à expliquer, repose sur l'utilisation des automates de THOMPSON ; elle avait été publiée en 1968.

Bien plus tôt, R. MCNAUGHTON et H. YAMADA (en 1960) et V. M. GLUSHKOV (en 1961) avaient, indépendamment, publié une méthode d'apparence très différente : elle consiste à numéroter les occurrences de chaque lettre de l'expression rationnelle, pour obtenir une nouvelle expression dite *linéaire*. La construction de l'automate fini associé est alors très simple ; on termine en «dénumérotant» les états et en déterminisant l'automate.

Gérard BERRY et Ravi SETHI ont donné en 1986 une présentation rigoureuse de cette méthode, et mis en évidence sa relation avec les *dérivées* d'une expression rationnelle (notion due à Janus BRZOZOWSKI).

Jean BERSTEL et Jean-Éric PIN ont proposé en 1995 une interprétation très simple de cette méthode, en faisant appel aux langages locaux.

Veuillez rédiger chaque partie sur une copie séparée.

Table des matières

Préliminaires : définitions et notations

▶ Soit L un langage sur l'alphabet X ; on note $\mathsf{Pref}(L)$ l'ensemble des préfixes de longueur 1 des mots de L, $\mathsf{Suff}(L)$ l'ensemble des suffixes de longueur 1 des mots de L et $\mathsf{Fact}(L)$ l'ensemble des facteurs de longueur 2 des mots de L. On a donc :

$$\begin{array}{rcl} x \in \mathsf{Pref}(L) & \iff & x \in X \text{ et } xX^* \cap L \neq \emptyset \\ x \in \mathsf{Suff}(L) & \iff & x \in X \text{ et } X^*x \cap L \neq \emptyset \\ u \in \mathsf{Fact}(L) & \iff & u \in X^2 \text{ et } X^*uX^* \cap L \neq \emptyset \end{array}$$

On définit également $\mathsf{Non}(L) = X^2 \setminus \mathsf{Fact}(L)$.

▶ Soit $\mathcal{A}$ un automate fini ; on note $\mathcal{L}_{AF}(\mathcal{A})$ le langage reconnu par $\mathcal{A}$. De même, si e est une expression rationnelle, on note $\mathcal{L}_{ER}(e)$ le langage décrit par e.

▶ Notre objectif est de décrire un algorithme calculant une fonction $\mathbf{A}$ qui, à une expression rationnelle e, associe un automate fini $\mathcal{A}$ tel que $\mathcal{L}_{AF}(\mathcal{A}) = \mathcal{L}_{ER}(e)$. On aura donc $\mathcal{L}_{AF} \circ \mathbf{A} = \mathcal{L}_{ER}$.

▶ Soient X et Y deux alphabets. Un *morphisme* de X^* vers Y^* est une application $\varphi : X^* \mapsto Y^*$ vérifiant $\varphi(uv) = \varphi(u)\varphi(v)$ quels que soient les mots u et v pris dans X^*. Il est clair qu'un tel morphisme est parfaitement défini par les images des éléments de X.

▶ Un automate fini déterministe $\mathcal{A} = (Q, \delta, i, F)$ est dit *standard* si aucune de ses transitions ne mène à l'état initial i.

Question 1 • Soit $\mathcal{A} = (Q, \delta, i, F)$ un automate fini déterministe. Décrivez un algorithme construisant un automate fini déterministe standard $\mathcal{A}'$ reconnaissant le même langage que $\mathcal{A}$. Nous dirons que $\mathcal{A}'$ est le *standardisé* de $\mathcal{A}$.

1 Langages locaux

▶ Un langage L sur l'alphabet X est dit *local* si, notant $\mathsf{P} = \mathsf{Pref}(L)$, $\mathsf{S} = \mathsf{Suff}(L)$, $\mathsf{N} = \mathsf{Non}(L)$ on a $L \setminus \{\varepsilon\} = (\mathsf{P}X^* \cap X^*\mathsf{S}) \setminus X^*\mathsf{N}X^*$. Ceci revient à dire qu'un mot appartient à L ssi sa première lettre est dans P, sa dernière lettre est dans S, et si aucun de ses facteurs de longueur 2 n'appartient à N.

Question 2 • Montrez que tout langage local est rationnel.

Question 3 • Montrez que le langage décrit par l'expression rationnelle $(abc)^*$ est local.

Question 4 • Montrez que le langage décrit par l'expression rationnelle a^*ba n'est pas local.

Question 5 • Exhibez deux langages locaux L_1 et L_2, dont la réunion $L_1 \cup L_2$ n'est pas un langage local.

Question 6 • Exhibez deux langages locaux L_1 et L_2, dont la concaténation $L_1 \cdot L_2$ n'est pas un langage local.

Question 7 • Montrez que l'image d'un langage rationnel par un morphisme est également un langage rationnel.

Question 8 • Soit $\mathcal{A} = (Q, \delta, I, F)$ un automate fini (déterministe ou non). Montrez que l'ensemble des calculs réussis de $\mathcal{A}$ est un langage local sur l'alphabet $Q \times X \times Q$.

▶ • Un morphisme $\varphi : X^* \mapsto Y^*$ est *alphabétique* si $|\varphi(x)| \leqslant 1$ pour toute lettre $x \in X$; il est *strictement alphabétique* si $|\varphi(x)| = 1$ pour toute lettre $x \in X$

Question 9 • Montrez que tout langage rationnel est l'image d'un langage local par un morphisme strictement alphabétique.

2 Automates locaux

▶ Un automate fini déterministe $\mathcal{A} = (Q, \delta, i, F)$ est dit *local* si les transitions étiquetées par une lettre x données arrivent toutes dans un même état, qui ne dépend donc que de x.

Question 10 • Montrez que tout langage local sur un alphabet X peut être reconnu par un automate local standard, dont l'ensemble des états est $X \cup \{\varepsilon\}$.

Question 11 • Réciproquement, montrez que le langage reconnu par un automate local est lui-même local.

Question 12 • Soient L_1 et L_2 deux langages locaux sur des alphabets X et Y disjoints. Montrez que $L_1 \cup L_2$ est un langage local.

Question 13 • Soient L_1 et L_2 deux langages locaux sur des alphabets X et Y disjoints. Montrez que $L_1 \cdot L_2$ est un langage local.

Question 14 • Soit L un langage local. Montrez que L^* est également un langage local.

3 L'algorithme de McNaughton, Yamada et Glushkov

► Une expression rationnelle e sur l'alphabet X est dite *linéaire* si chaque lettre de X possède au plus une occurrence dans e.

► Soit e une expression rationnelle. On obtient une *version linéaire* e' de e en remplaçant les lettres qui apparaissent dans e par des lettres deux à deux distinctes ; plus précisément, les diverses occurrences dans e d'une même mettre a seront remplacées par $a_1, a_2, \ldots, a_{|e|_a}$. Ainsi, une version linéaire de $e = (a+b)^*aba$ sera $e' = (a_1 + b_1)^* a_2 b_2 a_3$.

Question 15 • Soit e une expression rationnelle linéaire. Montrez que $\mathcal{L}_{ER}(e)$ est un langage local.

Question 16 • Que pensez-vous de la réciproque ?

Question 17 • Définissez, par induction structurelle, la fonction λ qui, à l'expression rationnelle e, associe $\{\varepsilon\} \cap \mathcal{L}_{ER}(e)$.

Question 18 • Définissez, toujours par induction structurelle, une fonction $\mathbf{P}$ qui, à l'expression rationnelle linéaire e, associe $\mathbf{P}(e) = \mathsf{Pref}\big(\mathcal{L}_{ER}(e)\big)$. Nous aurons donc $\mathbf{P} = \mathsf{Pref} \circ \mathcal{L}_{ER}$.

Question 19 • Donnez de même des définitions par induction structurelle des fonctions $\mathbf{S} = \mathsf{Suff} \circ \mathcal{L}_{ER}$ et $\mathbf{F} = \mathsf{Fact} \circ \mathcal{L}_{ER}$.

► L'algorithme de McNaughton-Yamada-Glushkov permet d'associer à une expression rationnelle e un automate fini reconnaissant le langage $\mathcal{L}_{ER}(e)$. Il consiste en quatre étapes :

1. construire une version linéaire e' de l'expression rationnelle e, en mémorisant l'encodage des lettres qui apparaissent dans e
2. construire les ensembles $\mathbf{P}(e')$, $\mathbf{S}(e')$ et $\mathbf{F}(e')$
3. construire un automate fini déterministe $\mathcal{A}'$ reconnaissant $\mathcal{L}_{ER}(e')$
4. décoder les étiquettes des transitions de $\mathcal{A}'$ pour obtenir un automate fini reconnaissant $\mathcal{L}_{ER}(e)$

Question 20 • Appliquez l'algorithme de McNaughton-Yamada-Glushkov à l'expression rationnelle $((ab(ac)^* + ca)^*b)^*$.

Question 21 • Expliquez la phrase suivante, qui conclut l'article dont est tiré ce texte :

> ≪Berry and Sethi have given an unusual proof of a well-known result, namely that every rational language is the homomorphic image of a local language.≫

FIN