Automates (3)

option informatique

Théorème de Kleene

Formulation du théorème

Le **théorème de Kleene** affirme l'équivalence entre langages rationnels et langages reconnaissables.

Théorème 1 (de Kleene)

Un langage L sur un alphabet Σ est rationnel si et seulement s'il existe un automate fini $\mathcal{A}$ tel que le langage reconnu $\mathcal{L}(\mathcal{A})$ par cet automate soit égal à L.

Démonstration

Pour établir que tout langage rationnel est reconnaissable par un automate fini :

- l'algorithme de Berry-Sethi décrit une procédure de construction d'un automate fini non déterministe qui reconnaît une expression rationnelle ;
- l'automate obtenu est appelé automate de Glushkov de l'expression rationnelle.

Le lemme d'Arden permet, à partir d'un automate, de construire une expression rationnelle qui dénote le langage reconnu par l'automate. C'est la réciproque du théorème.

  - arrtes@gmail.com

Langage rationnel ?

Pour montrer qu'un **langage** est **rationnel**, on peut :

- soit exhiber une expression rationnelle qui le dénote ;
- soit construire un automate qui le reconnaît.

On peut ajouter deux autres résultats importants.

- Les langages reconnus par un automate sont **clos par complémentation et par intersection**.
- Le **lemme de l'étoile** permet de montrer qu'un **langage** est **non rationnel**.

Clôture

Complémentation

Théorème 2 (Clôture par complémentation)

Les langages reconnaissables sont clos par complémentation.

Démonstration

- La démonstration s'appuie sur la notion d'**automate déterministe complet**.
- Soit L un langage reconnu par un automate déterministe complet $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$.
- Le **langage complémentaire** $\bar{L}$ de L est $\Sigma^* \setminus L$.
- Définissons l'automate déduit de $\mathcal{A}$ en remplaçant les états acceptants F par $Q \setminus F$. Alors, par construction :

$$\mathcal{A}' = (\Sigma, Q, q_0, Q \setminus F, \delta)$$

reconnaît $\bar{L}$.

Commentaires

- Le caractère **complet** est indispensable car un blocage de $\mathcal{A}$ serait aussi un blocage de $\mathcal{A}'$.
- L'idée de la démonstration est que les mots reconnus par $\mathcal{A}$ sont les mots qui se terminent sur un état acceptant de $\mathcal{A}$. Tous les mots de Σ^* qui se terminent sur un état non acceptant de $\mathcal{A}$ ne sont pas reconnus par $\mathcal{A}$. Ce mots sont ceux de $\Sigma^* \setminus L$. Ils sont reconnus par l'automate $\mathcal{A}'$ dont les états acceptant sont tous les états de $\mathcal{A}$ autres que ses états acceptants.

Intersection

Théorème 3 (Clôture par intersection)

Les langages reconnaissables sont clos par intersection.

Pour finir

- Nous venons de voir que les langages rationnels sont clos par **complémentation** et par **intersection**.
- Les propriétés de stabilité par **union**, par **concaténation** et par **passage à l'étoile de Kleene** des langages rationnels permettent d'en déduire que les langages reconnaissables sont clos par union, par concaténation et par passage à l'étoile de Kleene.
- Enfin, les langages reconnaissables sont clos par **passage au miroir**. Par le théorème de Kleene, les langages rationnels sont aussi stables par passage au miroir.

Lemme de l'étoile

Idée intuitive du lemme de l'étoile

- Le **lemme de l'étoile** affirme que passé une certaine taille, tout mot d'un langage L peut être construit par répétition d'un facteur itérant y s'insérant au sein d'un mot xz de L.
- Ce résultat, évident pour tout **langage fini**, reste valable pour tout **langage de cardinal infini**.
- Ce lemme est aussi nommé **lemme de pompage** car le facteur itérant y peut être « pompé » autant de fois que nécessaire pour produire un mot de L.
- Noter que le résultat n'est pas une équivalence mais une implication. Il existe des langages rationnels qui vérifient les conclusions du lemme.

Analyse

Théorème 4 (Lemme de l'étoile)

Si L est un langage rationnel, il existe $k \in \mathbb{N}$ tel que tout mot $m \in L$ vérifiant $|m| \geqslant k$ se factorise sous la forme $m = xyz$ avec $y \neq \varepsilon$, $|xy| \leqslant k$ et pour tout entier naturel n, $xy^nz \in L$.

*L'entier k, appelé la **constante d'itération**, ne dépend que de L.*
*y est appelé le **facteur itérant**.*

Démonstration

- Soit $\mathcal{A} = (\Sigma, Q, q_0, F, \delta)$ et $k = |Q|$.
- Soit m un mot de L tel que $|m| \geqslant k$.
- Le chemin $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \dots \xrightarrow{a_p} q_p$ reconnaissant m implique $p+1$ états.
- Il passe donc nécessairement deux fois par le même état $q_i = q_j$ avec $1 \leqslant i < j \leqslant k$.

$$q_0 \xrightarrow{a_1} \dots \xrightarrow{a_i} q_i \dots \xrightarrow{a_j} q_j \dots \xrightarrow{a_p} q_p$$

- Posons : $x = a_1 \dots a_i$ $y = a_{i+1} \dots a_j$ $z = a_{j+1} \dots a_p$.
  Alors, pour tout $n \in \mathbb{N}$, le chemin étiqueté par xy^nz conduit à l'état acceptant q_p.
- Donc $xy^nz \in L$.

Exemple 1

Exercice

Soit $\Sigma = \{a, b\}$ un alphabet. Considérons le langage défini par :

$$L = \{a^n b^n \mid n \in \mathbb{N}\}$$

Montrer que L n'est pas rationnel.

L'idée générale de la démonstration est d'établir **par l'absurde**, à l'aide du **lemme de l'étoile**, que le langage n'est pas **rationnel**.

Exemple 1 - démonstration 1

Démonstration 1

- Supposons L rationnel. Il existe $k \in \mathbb{N}^*$ tel que tout mot w vérifiant $|w| \geqslant k$ satisfait le **lemme de l'étoile**. Prenons $w = a^k b^k$.
- Puisque $|w| = 2k$, w est assez long et $w \in L$. D'après le **lemme de l'étoile**, il existe x, y, z tels que $w = xyz$, avec $|xy| \leqslant k$, $y \neq \varepsilon$, vérifiant :

$$\forall q \in \mathbb{N}, \quad xy^q z \in L$$

- Puisque la condition $|xy| \leqslant k$ doit être vérifiée, y doit apparaître dans les k premiers caractères de w sous la forme a^p, pour un certain entier p. Comme $y \neq \varepsilon$, alors : $p \geqslant 1$.
  Par exemple, $x = a^{k-p}$ et $y = a^p$.
- Écrivons alors xy^2 (lemme avec $q = 2$). Le mot résultant est de la forme :

$$a^{k+p} b^k$$

  Or, ce mot n'appartient pas à L alors que le lemme de l'étoile affirme le contraire.
- Il existe donc au moins un mot qui ne satisfait pas le lemme de l'étoile. ***L* n'est pas rationnel**.

Exemple 1

Démonstration 2

- Si L est reconnaissable, il existe un entier $n \geqslant 1$ tel que le lemme de l'étoile soit vérifié.
- Posons :

$$x = \varepsilon \qquad y = a^n \qquad z = b^n$$

- Alors, il existe $p \in \mathbb{N}^*$ tel que :

$$\forall k \in \mathbb{N} \qquad a^{n+kp} b^n \in L$$

- Ce résultat est absurde puisque de tels mots contiennent plus de a que de b. L **n'est pas rationnel**.

Exemple 2

Exercice

Soit $\Sigma = \{(,)\}$ un alphabet. Considérons le langage des expressions bien parenthésées défini par :

$$L = \{w \in \Sigma^* \mid w \text{ bien parenthésé}\}$$

Montrer que L n'est pas rationnel.

Exemple 2

Démonstration

- Remarquons d'abord que L contient des mots comme $(())((()())$ ou comme $(((())))$. Pour utiliser le **lemme de l'étoile**, l'idée est de considérer un mot aussi simple que possible. Par exemple, pour un entier naturel k :

$$w = (^k)^k$$

- Puisque $|w| = 2k$ et $w \in L$, w doit satisfaire les conditions du lemme de l'étoile. Il doit dont exister x, y, z tels que $w = xyz$, , avec $|xy| \leqslant k$, $y \neq \varepsilon$, vérifiant :

$$\forall q \in \mathbb{N}, \quad xy^q z \in L$$

- Puisque la condition $|xy| \leqslant k$ doit être vérifiée, y doit apparaître dans les k premiers caractères de w sous la forme $(^p$, pour un certain entier p. Comme $y \neq \varepsilon$, alors : $p \geqslant 1$.
- Puisque $q = 2$, alors $xy^q = (^{k+p}$. Puis :

$$xy^2 z = (^{k+p})^k$$

- Ce mot doit appartenir à L, en vertu du lemme de l'étoile. Or ce mot est une expression mal parenthésée, qui n'est pas dans L. Donc, ***L* n'est pas rationnel**.