Théorie des langages rationnels : THLR

CM 4

Uli Fahrenberg

EPITA Rennes

Septembre 2021

Aperçu

Programme du cours

1. Mots, langages
2. Langages rationnels, expressions rationnelles
3. Automates finis
4. Langages non-rationnels
5. Langages reconnaissables, minimisation

Hier : Automates finis

- poly chapitre 4, sections 4.1.1, 4.1.2, 4.1.4 première moitié, 4.1.5
- plus la moitié de section 4.2.2

Hier : Automates finis

Définition

Un automate fini (à transitions spontanées) : $A = (\Sigma, Q, Q_0, F, \delta)$:

- Σ : ensemble fini de symboles, Q : ensemble fini d'états
- $Q_0 \subseteq Q$: états initiaux, $F \subseteq Q$: états finaux
- $\delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$: relation de transition

- on note $q \xrightarrow{a} r$ si $(q, a, r) \in \delta$

Définition (Sémantique de A)

- Un calcul dans A : $\sigma = q_1 \xrightarrow{a_1} q_2 \xrightarrow{a_2} \cdots \xrightarrow{a_{n-1}} q_n$
- L'étiquette d'un calcul : $\lambda(\sigma) = a_1 a_2 \dots a_{n-1} \in \Sigma^*$
- Un calcul réussi : $q_1 = q_0$ et $q_n \in F$
- Le langage reconnu par A :
  $L(A) = \{\lambda(\sigma) \mid \sigma \text{ calcul réussi dans } A\}$

Variants

Un automate fini (à transitions spontanées) : $A = (\Sigma, Q, q_0, F, \delta)$:

- Σ, Q ensembles finis, $q_0 \in Q$, $F \subseteq Q$,
- $\delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$: la relation de transition

Définition

- A est sans transitions spontanées si $\delta \subseteq Q \times \Sigma \times Q$.
- A est complet si $\forall q \in Q : \forall a \in \Sigma : |\{r \in Q \mid (q, a, r) \in \delta\}| \geq 1$.
- A est déterministe si $\delta \subseteq Q \times \Sigma \times Q$, $|Q_0| = 1$ et
  $\forall q \in Q : \forall a \in \Sigma : |\{r \in Q \mid (q, a, r) \in \delta\} \leq 1$.

Hier dans l'ordre

1. automates finis déterministes complets
2. automates finis déterministes
3. automates finis (sans transitions spontanées)
4. automates finis (à transitions spontanées)

Variants

Un automate fini (à transitions spontanées) : $A = (\Sigma, Q, q_0, F, \delta)$:

- Σ, Q ensembles finis, $q_0 \in Q$, $F \subseteq Q$,
- $\delta \subseteq Q \times (\Sigma \cup \{\varepsilon\}) \times Q$: la relation de transition

Définition

- A est sans transitions spontanées si $\delta \subseteq Q \times \Sigma \times Q$.
- A est complet si $\forall q \in Q : \forall a \in \Sigma : |\{r \in Q \mid (q, a, r) \in \delta\}| \geq 1$.
- A est déterministe si $\delta \subseteq Q \times \Sigma \times Q$, $|Q_0| = 1$ et $\forall q \in Q : \forall a \in \Sigma : |\{r \in Q \mid (q, a, r) \in \delta\} \leq 1$.

Hier dans l'ordre

1. automates finis déterministes complets DFA
2. automates finis déterministes
3. automates finis (sans transitions spontanées) NFA
4. automates finis (à transitions spontanées) ε-NFA

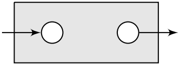
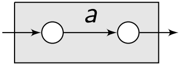
Hier : Langages reconnaissables

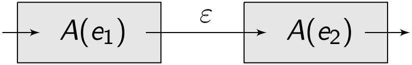
Définition

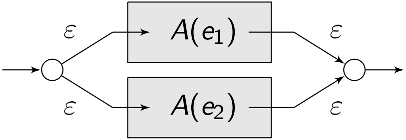
Un langage L est reconnaissable si $\exists$ un automate fini A t.q. $L = L(A)$.

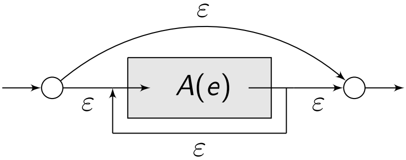
syntaxe		sémantique
aut. finis dét. complets		langages reconnaissables
$\not\supseteq\!\cap$		$\|\|$ ✓
aut. finis déterministes		langages reconnaissables
$\not\supseteq\!\cap$	$\xrightarrow{L(\cdot)}$	$\|\|$?
automates finis		langages reconnaissables
$\not\supseteq\!\cap$		$\|\|$ ✓
aut. finis à trans. spontanées		langages reconnaissables
		$\|\|$ $\Uparrow$
expressions rationnelles		langages rationnelles

Hier : Algorithme de Thompson

- pour traduire une expression rationnelle e en automate fini $A(e)$, inductivement
- $\emptyset \mapsto$  ; $a \in \Sigma \cup \{\varepsilon\} \mapsto$ 

- $e_1 e_2 \mapsto$ 

- $e_1 + e_2 \mapsto$ 

- $e^* \mapsto$ 

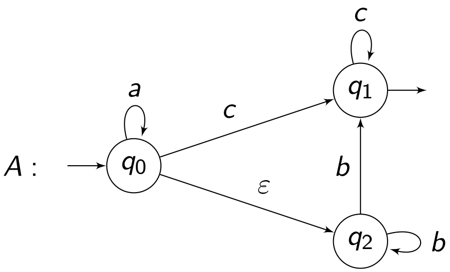
Quelques retours QCM 2

1. Un langage quelconque
   - contient toujours un langage rationnel.
   - est toujours dénombrable.
2. Quel mot de la liste ci-dessous appartient au langage dénoté par l'expression `[a-zA-Z_]*[0-9]+_(-)?[0-9]+`?
   - `_aZza1_0`
3. Quel mot de la liste ci-dessous appartient au langage dénoté par l'expression `[0-9]*[a-z]1_(-)?[0-9]+`?
   - `a1_-0`
4. Pour toute expression régulière e, on a $\emptyset e \equiv e \emptyset \equiv \emptyset$. ✓
5. Pour toute expression régulière e, on a $e^* \equiv (e^*)^*$. ✓
6. Pour toute expression régulière e, on a $ee \equiv e$. ✗
7. Pour toute expression régulière e, on a $\emptyset + e \equiv e + \emptyset \equiv e$. ✓

Aujourd'hui

1. Convertir un automate fini en automate fini déterministe : déterminisation
2. Méthode pour démontrer qu'il existe des langages non-rationnels

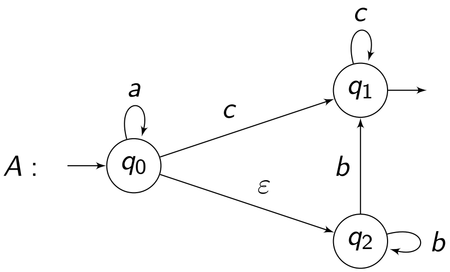
5 minutes de réflexion



Vrai ou faux ?

1. $acc \in L(A)$
2. $acb \in L(A)$
3. $abc \in L(A)$
4. $abb \in L(A)$
5. $L(b^*bc^*) \subseteq L(A)$

5 minutes de réflexion



Vrai ou faux ?

1. $acc \in L(A)$ ✓
2. $acb \in L(A)$ ✗
3. $abc \in L(A)$ ✓
4. $abb \in L(A)$ ✓
5. $L(b^* bc^*) \subseteq L(A)$ ✓

Déterminisation

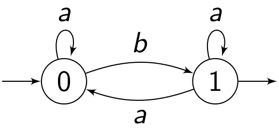
Automate des parties

Définition

Soit $A = (\Sigma, Q, Q_0, F, \delta)$ un automate fini. L'automate des parties de A est l'automate fini déterministe complet $A' = (\Sigma, Q', q_0', F', \delta')$ définit comme suite :

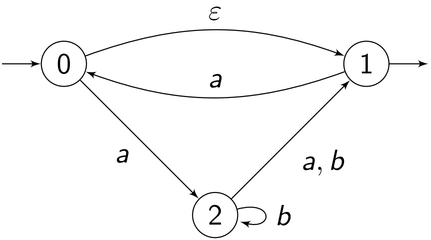
- $Q' = \mathcal{P}(Q)$, l'ensemble des parties de Q,
- $q_0' = Q_0$,
- $F' = \{P \subseteq Q \mid P \cap F \neq \emptyset\}$, et
- $\delta'(P, a) = \{q \in Q \mid \exists p \in P : (p, a, q) \in \delta\}$.

Exemple (sur tableau)



Exemple (sur tableau)

- et ça marche aussi avec transitions spontanées :



Le non-déterminisme ne paye pas

Théorème

Pour tout automate fini A il existe un automate fini déterministe complet A' tel que $L(A') = L(A)$.

Démonstration.

1. Notons $A = (\Sigma, Q, Q_0, F, \delta)$.
2. Soit A' l'automate des parties de A, on montre que $L(A') = L(A)$.

Le non-déterminisme ne paye pas

Théorème

Pour tout automate fini A il existe un automate fini déterministe complet A' tel que $L(A') = L(A)$.

Démonstration.

1. Notons $A = (\Sigma, Q, Q_0, F, \delta)$.
2. Soit A' l'automate des parties de A, on montre que $L(A') = L(A)$.
3. Soit $w \in L(A)$, alors il existe un calcul réussi $\sigma = q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} q_n$ dans A t.q. $\lambda(\sigma) = w$.

Le non-déterminisme ne paye pas

Théorème

Pour tout automate fini A il existe un automate fini déterministe complet A' tel que $L(A') = L(A)$.

Démonstration.

1. Notons $A = (\Sigma, Q, Q_0, F, \delta)$.
2. Soit A' l'automate des parties de A, on montre que $L(A') = L(A)$.
3. Soit $w \in L(A)$, alors il existe un calcul réussi $\sigma = q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} q_n$ dans A t.q. $\lambda(\sigma) = w$.
4. Soit $Q_1 = \delta'(Q_0, a)$, $Q_2 = \delta'(Q_1)$ etc., alors $\sigma' = Q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} Q_n$ est un calcul dans A' t.q. $\lambda(\sigma') = w$.

Le non-déterminisme ne paye pas

Théorème

Pour tout automate fini A il existe un automate fini déterministe complet A' tel que $L(A') = L(A)$.

Démonstration.

1. Notons $A = (\Sigma, Q, Q_0, F, \delta)$.
2. Soit A' l'automate des parties de A, on montre que $L(A') = L(A)$.
3. Soit $w \in L(A)$, alors il existe un calcul réussi $\sigma = q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} q_n$ dans A t.q. $\lambda(\sigma) = w$.
4. Soit $Q_1 = \delta'(Q_0, a)$, $Q_2 = \delta'(Q_1)$ etc., alors $\sigma' = Q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} Q_n$ est un calcul dans A' t.q. $\lambda(\sigma') = w$.
5. On a $q_i \in Q_i$ pour tout i, donc $q_n \in Q_n \cap F$, c.à.d. $Q_n \in F'$, alors σ' est un calcul réussi, donc $w \in L(A')$.

Le non-déterminisme ne paye pas

Théorème

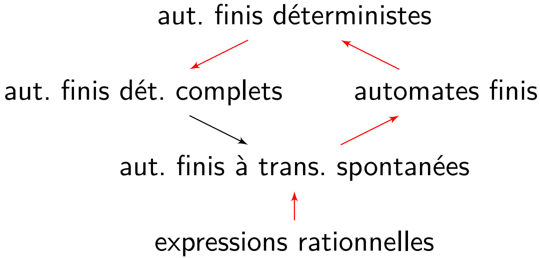
Pour tout automate fini A il existe un automate fini déterministe complet A' tel que $L(A') = L(A)$.

Démonstration.

1. Notons $A = (\Sigma, Q, Q_0, F, \delta)$.
2. Soit A' l'automate des parties de A, on montre que $L(A') = L(A)$.
3. Soit $w \in L(A)$, alors il existe un calcul réussi $\sigma = q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} q_n$ dans A t.q. $\lambda(\sigma) = w$.
4. Soit $Q_1 = \delta'(Q_0, a)$, $Q_2 = \delta'(Q_1)$ etc., alors $\sigma' = Q_0 \xrightarrow{a_1} \cdots \xrightarrow{a_n} Q_n$ est un calcul dans A' t.q. $\lambda(\sigma') = w$.
5. On a $q_i \in Q_i$ pour tout i, donc $q_n \in Q_n \cap F$, c.à.d. $Q_n \in F'$, alors σ' est un calcul réussi, donc $w \in L(A')$.

Et l'autre direction ?

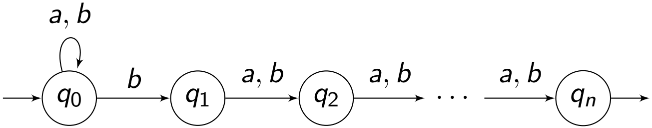
Le non-déterminisme paye



- difficile d'inventer un traduction directe des expressions rationnelles en automates finis déterministes
- le non-déterminisme est utile pour des spécifications partielles
- des automates finis non-déterministes peuvent être exponentiellement plus distinctes que des automates finis déterministes :

Exercice

Pour $n \geq 2$ soit A_n l'automate fini comme suit :

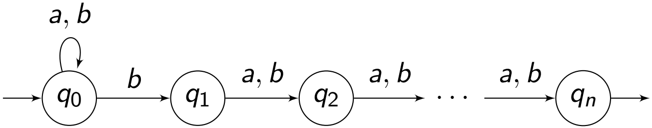


1. Trouver une expression rationnelle e_n telle que $L(e_n) = L(A_n)$.

2. Quelle est le nombre d'états le plus petit d'un automate fini déterministe A'_n tel que $L(A'_n) = L(A_n)$?

Exercice

Pour $n \geq 2$ soit A_n l'automate fini comme suit :



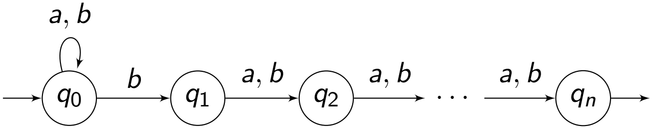
1. Trouver une expression rationnelle e_n telle que $L(e_n) = L(A_n)$.

   $(a + b)^* b(a + b)^{n-1}$

2. Quelle est le nombre d'états le plus petit d'un automate fini déterministe A'_n tel que $L(A'_n) = L(A_n)$?

Exercice

Pour $n \geq 2$ soit A_n l'automate fini comme suit :



1. Trouver une expression rationnelle e_n telle que $L(e_n) = L(A_n)$.

   $(a+b)^* b (a+b)^{n-1}$

2. Quelle est le nombre d'états le plus petit d'un automate fini déterministe A'_n tel que $L(A'_n) = L(A_n)$?

   2^n

Langages non-rationnels

Motifs répétitifs

- Existent-ils des langages non-rationnels ?
- Le langage $\{a^n b^n \mid n \geq 0\}$ est-il rationnel ?
- Le langage des expressions arithmétiques est-il rationnel ?

Lemme de l'étoile

1. Soit $A = (\Sigma, Q, Q_0, F, \delta)$ un automate fini avec k états.
2. Soit $x \in L(A)$ un mot de longueur $|x| = k$ (si il existe) ; écrivons $x = a_1 \dots a_k$.
3. Alors on a un calcul réussi $s_1 \xrightarrow{a_1} s_2 \xrightarrow{a_2} \cdots \xrightarrow{a_k} s_{k+1}$ dans A.
4. Ce calcul utilise $k+1$ états, alors un état de A a été utilisé deux fois. (Principe des tiroirs.)
5. Soient donc $i < j$ tel que $s_i = s_j$: la chaîne $s_i \rightsquigarrow s_j$ est une boucle.
6. Alors $s_1 \xrightarrow{a_1} \cdots \xrightarrow{a_{i-1}} s_i \xrightarrow{a_j} s_{j+1} \to \cdots \to s_{k+1}$ est aussi un calcul réussi, avec étiquette $a_1 \dots a_{i-1} a_j \dots a_k$.
7. En écrivant $u = a_1 \dots a_{i-1}$, $v = a_i \dots a_{j-1}$ et $w = a_j \dots a_k$ on trouve que $L(uv^*w) \subseteq L(A)$.

Lemme de l'étoile

Théorème (4.25)

Soit L un langage rationnel. Il existe $k \geq 0$ tel que tout $x \in L$ avec longueur $|x| \geq k$ peut s'écrire $x = uvw$ avec $|uv| \leq k$, $|v| \geq 1$ et $L(uv^*w) \subseteq L$.

- aussi lemme de pompage
- note $\exists k : \forall x : \exists u, v, w$
- démonstration par quelques petites modifications de l'argument précédent

Corollaire

Théorème (rappel)

Soit L un langage rationnel. Il existe $k \geq 0$ tel que tout $x \in L$ avec longueur $|x| \geq k$ peut s'écrire $x = uvw$ avec $|uv| \leq k$, $|v| \geq 1$ et $L(uv^*w) \subseteq L$.

Corollaire

Le langage $\{a^n b^n \mid n \geq 0\}$ n'est pas rationnel.

Démonstration.

1. Supposons par l'absurde que L soit rationnel.
2. Soit k comme fourni par la lemme d'étoile.
3. Soit $x = a^k b^k$, alors $x = uvw$ avec $|uv| \leq k$ et $|v| \geq 1$.
4. Donc $u = a^i$, $v = a^j$ et $w = a^{k-i-j} b^k$ pour un $j \geq 1$.
5. On a $uw \in L(uv^*w)$ mais $uw \notin L$, contradiction !

Exercice

Théorème (rappel)

Soit L un langage rationnel. Il existe $k \geq 0$ tel que tout $x \in L$ avec longueur $|x| \geq k$ peut s'écrire $x = uvw$ avec $|uv| \leq k$, $|v| \geq 1$ et $L(uv^*w) \subseteq L$.

Montrer que le langage $\{ww \mid w \in \{a,b\}^*\}$ n'est pas rationnel.

Les automates finis sont décidables

Théorème (4.27)

Il existe un algorithme qui, pour A un automate fini, décide si $L(A)$ est vide, fini ou infini.

Démonstration.

Soit k le nombre d'états de A.

1. $L(A)$ est non-vide ssi il existe $w \in L(A)$ avec longueur $|w| < k$.
2. $L(A)$ est infini ssi il existe $w \in L(A)$ avec $k \leq |w| < 2k$.

(le reste sur tableau)


That's all Folks!