Langages formels : expressions rationnelles

Stéphane Le Roux stephane.le_roux@ens-paris-saclay.fr

ENS Paris-Saclay

2023-2024

Expressions rationnelles

Définition

Soit Σ un alphabet fini et non-vide. Les expressions rationnelles sur Σ sont des mots sur l'alphabet $\underline{\Sigma} \cup \{(,), +, \cdot,^*, \underline{\emptyset}\}$. Leur ensemble $\mathrm{ER}(\Sigma)$ est défini par récurrence structurelle.

- $\underline{\emptyset} \in \mathrm{ER}(\Sigma)$
- Pour tout $a \in \Sigma$, on pose $\underline{a} \in \mathrm{ER}(\Sigma)$
- Pour tout $E, F \in \mathrm{ER}(\Sigma)$ on pose
  - $(E + F) \in \mathrm{ER}(\Sigma)$
  - $(E \cdot F) \in \mathrm{ER}(\Sigma)$
  - $(E^*) \in \mathrm{ER}(\Sigma)$

Sémantique des expressions rationnelles

À chaque expression rationnelle sur Σ, on associe un langage sur Σ, par récurrence structurelle.

- $\mathcal{L}(\underline{\emptyset}) := \emptyset$
- Pour tout $a \in \Sigma$, on pose $\mathcal{L}(\underline{a}) := \{a\}$
- Pour tout $E, F \in \mathrm{ER}(\Sigma)$ on pose
  - $\mathcal{L}(E + F) := \mathcal{L}(E) \cup \mathcal{L}(F)$
  - $\mathcal{L}(E \cdot F) := \mathcal{L}(E) \cdot \mathcal{L}(F)$
  - $\mathcal{L}(E^*) := \mathcal{L}(E)^*$

Lemme

- $\mathcal{L}(E + F) = \mathcal{L}(F + E)$
- $\mathcal{L}((E + F) + G) = \mathcal{L}(E + (F + G))$
- $\mathcal{L}((E \cdot F) \cdot G) = \mathcal{L}(E \cdot (F \cdot G))$
- $\mathcal{L}(E \cdot (F + G)) = \mathcal{L}((E \cdot F) + (E \cdot G))$
- $\mathcal{L}((E + F) \cdot G) = \mathcal{L}((E \cdot G) + (F \cdot G))$
- $\mathcal{L}((E^*)^*) = \mathcal{L}(E^*)$
- $\mathcal{L}(\underline{\emptyset}^*) = \{\epsilon\}$

Abus de notation

- On omet souvent de souligner les lettres dans les expressions rationnelles.
- On omet souvent les parenthèses externes ou rendues inutiles par associativité.
- On écrit parfois $E + F \cdot G$ au lieu de $E + (F \cdot G)$.
- On confond souvent expression rationnelle et langage associé.
- On écrit souvent Σ au lieu de $a + b + \dots$.

Expressions rationnelles et langages rationnels

Rappel :

- $\mathcal{L}(\underline{\emptyset}) := \emptyset$
- Pour tout $a \in \Sigma$, on pose $\mathcal{L}(\underline{a}) := \{a\}$
- Pour tout $E, F \in \mathrm{ER}(\Sigma)$ on pose
  - $\mathcal{L}(E + F) := \mathcal{L}(E) \cup \mathcal{L}(F)$
  - $\mathcal{L}(E \cdot F) := \mathcal{L}(E) \cdot \mathcal{L}(F)$
  - $\mathcal{L}(E^*) := \mathcal{L}(E)^*$

Lemme

$\mathcal{L}[\mathrm{ER}(\Sigma)]$ est le $+$ petit ensemble de langages sur Σ contenant le langage vide et les $\{a\}$, et qui soit clos par union, concaténation et itération.

Preuve : par propriété des définitions par récurrence structurelle.

- On va montrer $\mathcal{L}[\mathrm{ER}(\Sigma)] = \mathrm{Rec}(\Sigma^*)$, en invoquant les propriétés de clôture de $\mathcal{L}[\mathrm{ER}(\Sigma)]$ et de $\mathrm{Rec}(\Sigma^*)$. Ainsi, on ne pourrait pas prouver les clôtures de $\mathrm{Rec}(\Sigma^*)$ en invoquant celles de $\mathcal{L}[\mathrm{ER}(\Sigma)]$.
- Je ne connais pas de preuve directe de clôture de $\mathcal{L}[\mathrm{ER}(\Sigma)]$ par intersection ou complémentaire.

Expressions rationnelles et langages rationnels (II)

Lemme (Rappel)

$\mathcal{L}[\mathrm{ER}(\Sigma)]$ est le $+$ petit ensemble de langages sur Σ contenant le langage vide et les $\{a\}$, et qui soit clos par union, concaténation et itération.

Lemme

$\mathcal{L}[\mathrm{ER}(\Sigma)] \subseteq \mathrm{Rec}(\Sigma^*)$.

Preuve

$\mathrm{Rec}(\Sigma^*)$ contient le langage vide et les $\{a\}$, et est clos par union, concaténation et itération.

Expressions rationnelles vers langages rationnels

On définit une application des expressions rationnelles vers les ϵ-AFN.

- $g(\underline{\emptyset})$ est un ϵ-AFN à un état, non acceptant.
- $g(\underline{a})$ est un ϵ-AFN acceptant seulement le mot a.
- $g(E + F) := Union(g(E), g(F))$
- $g(E \cdot F) := Concat(g(E), g(F))$
- $g(E^*) := Iter(g(E))$

Lemme

1. Pour tout $E \in \mathrm{ER}(\Sigma)$ on a $\mathcal{L}(E) = \mathcal{L}(g(E))$
2. L'application g est calculable. (Logspace)

1. Par récurrence structurelle sur E.
2. Construction/modification des automates vue en cours.

On pourrait définir g' vers les AFD : toujours calculable mais explosion exponentielle.

Expressions rationnelles vers langages rationnels (II)

équivalence entre expressions rationnelles

Pour tout $E, F \in \mathrm{ER}(\Sigma)$. on note $E \equiv F$ si $\mathcal{L}(E) = \mathcal{L}(F)$.

Lemme

Le problème suivant est décidable.

- Entrée : Σ et $E, F \in \mathrm{ER}(\Sigma)$
- Sortie : Oui si $E \equiv F$

Preuve

On calcule $g(E)$ et $g(F)$, puis on décide si $\mathcal{L}(g(E)) = \mathcal{L}(g(F))$.

Langages rationnels vers expressions rationnelles

Algorithme de McNaughton-Yamada

Soit $\mathcal{A} = (\Sigma, [n], \Delta, I, F)$ un AFN, où $[n] = \{1, \ldots, n\}$.

Langages intermédaires

- $L_{p,q} := \mathcal{L}(\mathcal{A}[I \leftarrow \{p\}, F \leftarrow \{q\}])$
- Soit $L^k_{p,q}$ le langage des mots non vides qui, dans $\mathcal{A}$, amènent de p à q en ne visitant que des états intermédiaires de $[k]$. Formellement, $L^k_{p,q} := \{u \in \Sigma^+ \mid \delta^*(p, u) = q \wedge \forall i \in [|u| - 1], \delta(p, u_1 \ldots u_i) \in [k]\}$

Lemme

- $\mathcal{L}(\mathcal{A}) = \cup_{p \in I, q \in F} L_{p,q}$
- Si $p \neq q$ alors $L_{p,q} = L^n_{p,q}$
- $L_{p,p} = L^n_{p,p} \cup \{\epsilon\}$
- $L^0_{p,q} = \{a \in \Sigma \mid p \xrightarrow{a}_{\mathcal{A}} q\}$
- Pour tout $k > 0$ on a $L^k_{p,q} = L^{k-1}_{p,q} + L^{k-1}_{p,k} \cdot (L^{k-1}_{k,k})^* \cdot L^{k-1}_{k,q}$

Langages rationnels vers expressions rationnelles

Algorithme de McNaughton-Yamada

Soit $\mathcal{A} = (\Sigma, [n], \Delta, I, F)$ un AFN, où $[n] = \{1, \ldots, n\}$.

Lemme (rappel)

- $L^0_{p,q} = \{a \in \Sigma \mid p \xrightarrow{a}_{\mathcal{A}} q\}$
- Pour tout $k > 0$ on a $L^k_{p,q} = L^{k-1}_{p,q} + L^{k-1}_{p,k} \cdot (L^{k-1}_{k,k})^* \cdot L^{k-1}_{k,q}$

Lemme

On construit des expressions rationnelles par récurrence sur $k \in [n]$.

- Pour tout $p, q \in [n]$, soit $E^0_{p,q} := +\{a \in \Sigma \mid p \xrightarrow{a}_{\mathcal{A}} q\}$
- Pour $k, p, q \in [n]$, on pose $E^k_{p,q} := E^{k-1}_{p,q} + E^{k-1}_{p,k} \cdot (E^{k-1}_{k,k})^* \cdot E^{k-1}_{k,q}$

Alors pour tout $p, q, k \in [n]$ on a $\mathcal{L}(E^k_{p,q}) = \mathcal{L}^k_{p,q}$

Preuve

Par récurrence sur k.

Langages rationnels vers expressions rationnelles

Algorithme de McNaughton-Yamada

Soit $\mathcal{A} = (\Sigma, [n], \Delta, I, F)$ un AFN, où $[n] = \{1, \ldots, n\}$.

Lemme (rappel)

- $\mathcal{L}(\mathcal{A}) = \cup_{p \in I, q \in F} L_{p,q}$
- Si $p \neq q$ alors $L_{p,q} = L^n_{p,q}$
- $L_{p,p} = L^n_{p,p} \cup \{\epsilon\}$
- pour tout $p, q, k \in [n]$ on a $\mathcal{L}(E^k_{p,q}) = \mathcal{L}^k_{p,q}$

Lemme

1. Si $p \neq q$ soit $E_{p,q} := E^n_{p,q}$
2. Soit $E_{p,p} := E^n_{p,p} + \underline{\emptyset}^*$
3. Soit $E := +_{p \in I, q \in F} E_{p,q}$

Alors $\mathcal{L}(E) = \mathcal{L}(\mathcal{A})$

Langages rationnels et expressions rationnelles

Théroème (Kleene)

$\mathrm{Rec}(\Sigma^*) = \mathcal{L}(\mathrm{ER}(\Sigma))$

Preuve

- $\mathrm{Rec}(\Sigma^*) \subseteq \mathcal{L}(\mathrm{ER}(\Sigma))$ par le transparent précédent.
- $\mathcal{L}(\mathrm{ER}(\Sigma)) \subseteq \mathrm{Rec}(\Sigma^*)$ montré un peu avant.

Résumé

Les quatre assertions suivantes sont équivalentes

- L est reconnu par un AFD (ou AFN, etc)
- L a un nombre fini de résiduels
- L est produit par une expression rationnelle.
- $\exists N \in \mathbb{N}\setminus\{0\}, \forall (v, u_1, \ldots, u_N, w) \in \Sigma^* \times (\Sigma^+)^N \times \Sigma^*, \exists 0 < j < k \leq N$ tels que $vu_1 \ldots u_N w \in L$ ssi $vu_1 \ldots u_j u_{k+1} \ldots u_N w \in L$ (+ variantes)