Langages formels : introduction

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

ENS Paris-Saclay

2023-2024

Organisation

- 1ère partie (7 cours) par moi sur les automates, 2ème partie par Stefan Schwoon sur les automates à pile + TP. TD par Isa Vialard.
- Évaluation : partiel $\frac{1}{4}$, examen $\frac{1}{4}$, contrôle continu 1ère partie $\frac{1}{6}$, contrôle continu 2ème partie $\frac{1}{6}$, TP $\frac{1}{6}$.

Contexte

- Étant donné un alphabet $\Sigma \neq \emptyset$, on note Σ^* l'ensemble des mots finis écrits avec les lettres de Σ. Un langage est un sous-ensemble de Σ^*.
- Si Σ est dénombrable (fini ou infini), Σ^* est infini dénombrable et l'ensemble des langages $L \subseteq \Sigma^*$ est indénombrable.
- L'ensemble des langages reconnaissables par machines de Turing est dénombrable.
- Des modèles de calcul qui calculent les mêmes fonctions ou reconnaissent les mêmes langages : machines de Turing, fonctions récursives, lambda termes, grammaires générales, etc.
- On considère dans ce cours des modèles de calcul moins puissants et/ou des classes de langages plus simples, mais avec de meilleures propriétés.
- Un robot envoyé sur mars a une mémoire finie fixe car les extensions sont difficiles.

Les langages rationnels

Ils peuvent se définir de diverses manières équivalentes.

1. L a un nombre fini de résiduels $u^{-1}L$.
2. L est produit par une expressions rationnelles.
3. L est produit par une grammaire rationnelles.
4. L est reconnu par un automate fini.
5. L est reconnu par un morphisme de monoïdes.
6. La congruence syntaxique de L a un nombre fini de classes.
7. L est exprimable en logique monadique du second ordre.

Ces caractérisations fournissent perspectives et techniques de preuve variées.

On étudiera aussi

- une sous-classe des langage rationnels: les langages sans étoile,
- des grammaires engendrant des sur-classes diverses.

Perspective algébrique : moins efficace pour les modèles plus puissants.

Préliminaires

Qu'est-ce qu'un mot ?

Soit Σ un ensemble fini non vide qu'on appelle alphabet. Un mot peut être vu comme une liste d'éléments de Σ, qu'on appelle les lettres.

- ϵ (le mot vide) est un mot.
- Si u est un mot et $a \in \Sigma$, alors $a :: u$ est un mot.

On prouve $\forall u \in \Sigma^*, P(u)$ par récurrence structurelle en prouvant que :

- $P(\epsilon)$ est vraie, et
- $\forall u \in \Sigma^*, \forall a \in \Sigma, P(u) \Rightarrow P(a :: u)$

On peut définir la longueur $|\cdot| : \Sigma^* \to \mathbb{N}$ par récurrence structurelle, permettant les preuves par récurrence sur la longueur.

- $|\epsilon| := 0$
- $|a :: u| := |u| + 1$

Un mot peut aussi être vu comme une application d'un ordinal fini vers Σ. Par exemple *abac* est une application de $\{0, 1, 2, 3\}$ vers $\{a, b, c\}$.

Si $a \in \Sigma$, alors a représente aussi le mot constitué de l'unique lettre a.

Préfixe, suffixe, facteur, sous-mots

Les notions de préfixe et suffixe d'un mot se définissent bien par récurrence, mais de manières non symétriques.

- $\mathit{suff}(u, u)$
- $\mathit{suff}(u, v) \Rightarrow \mathit{suff}(u, a :: v)$ (récurrence structurelle)
- $\mathit{pref}(\epsilon, v)$ (aussi noté $\epsilon \sqsubseteq v$)
- $\mathit{pref}(u, v) \Rightarrow \mathit{pref}(a :: u, a :: v)$ (récurrence bien fondée sur le produit)

Les facteurs se définissent bien en utilisant les préfixes et en modifiant la définition pour les suffixes.

- $\mathit{pref}(u, v) \Rightarrow \mathit{fact}(u, v)$
- $\mathit{fact}(u, v) \Rightarrow \mathit{fact}(u, a :: v)$

Et avec la notion de suffixe ?

Les sous-mots s'obtiennent en modifiant la définition pour les préfixes.

- $\mathit{sousmot}(\epsilon, v)$
- $\mathit{sousmot}(u, v) \Rightarrow \mathit{sousmot}(u, a :: v) \wedge \mathit{sousmot}(a :: u, a :: v)$

Et avec la notion de suffixe ?

Préfixe, suffixe, facteur, sous-mots (II)

On peut définir la concaténation de $u, v \in \Sigma^*$ par récurrence structurelle sur u. (Et sur v ?)

- $\epsilon \cdot v := v$
- $(a :: u) \cdot v := a :: (u \cdot v)$

Caractérisations des notions de préfixe, suffixe et facteur.

- $u \sqsubseteq v$ ssi $\exists w \in \Sigma^*, u \cdot w = v$
- $\mathit{suff}(u, v)$ ssi $\exists w \in \Sigma^*, w \cdot u = v$
- $\mathit{fact}(u, v)$ ssi $\exists w_1, w_2 \in \Sigma^*, w_1 \cdot u \cdot w_2 = v$.

Remarque :

- Quand mot est vu comme une application d'un ordinal fini vers Σ, la concaténation se définit bien.
- Les notions de préfixes, suffixes, facteurs se définissent via les caractérisations ci-dessus.
- Le principe de preuve récurrence sur l'ordinal fonctionne, et permet de définir la notion de sous-mot.

Préfixe, suffixe, facteur, sous-mots (III)

La notion de miroir d'un mot se défini par récurrence structurelle et au moyen de la concaténation.

- $\tilde{\epsilon} := \epsilon$
- $\widetilde{a :: u} := \tilde{u} \cdot a$

Lemma

1. $\tilde{\tilde{u}} = u$, *par récurrence structurelle sur* u.
2. $\widetilde{uv} = \tilde{v}\tilde{u}$, *par récurrence structurelle sur* u *ou bien sur* v.
3. $\mathit{pref}(u, v)$ *ssi* $\mathit{suff}(\tilde{u}, \tilde{v})$

Lemma

$(\Sigma^*, \cdot, \epsilon)$ *est un monoïde, i.e. on peut prouver que* ϵ *est élément neutre et que* $\cdot$ *est associative.*

Récurrences

Rappel : on prouve $\forall u \in \Sigma^*, P(u)$ par récurrence structurelle en prouvant que :

- $P(\epsilon)$ est vraie, et
- $\forall u \in \Sigma^*, \forall a \in \Sigma, P(u) \Rightarrow P(a :: u)$

On peut aussi prouver que si

- $P(\epsilon)$ est vraie, et
- $\forall u \in \Sigma^*, \forall a \in \Sigma, P(u) \Rightarrow P(ua)$

alors $\forall u \in \Sigma^*, P(u)$.

Opérations sur les langages

- Opérations ensemblistes : union, intersection, différence, etc.
- Concaténation $L \cdot K := \{u \cdot v \mid (u, v) \in L \times K\}$.
- $L^0 := \{\epsilon\}$ et $L^{n+1} := L \cdot L^n = L^k \cdot L^{n+1-k}$.
- $L^* := \cup_{n \in \mathbb{N}} L^n$ et $L^+ := \cup_{n \in \mathbb{N} \setminus \{0\}} L^n$
- Définition par récurrence structurelle, clôture transitive (et réflexive)
- $\tilde{L} := \{\tilde{u} \mid u \in L\}$

Lemma

- $\widetilde{L \cdot K} = \tilde{K} \cdot \tilde{L}$.
- $(\mathcal{P}(\Sigma^*), \cdot, \{\epsilon\})$ *est un monoïde.*
- $(\mathcal{P}(\Sigma^*), \cup, \emptyset)$ *est un monoïde.*
- *La concaténation est distributive par rapport à l'union.*
- $L \cdot \emptyset = \emptyset \cdot L = \emptyset$.
- *Si* $K, L \neq \emptyset$ *alors* $|K| + |L| - 1 \overset{?}{\leq} |K \cdot L| \leq |K||L| = |K \times L|$.

Opérations sur les langages (II)

Pour tout $L, K \subseteq \Sigma^*$ on pose :

- $K^{-1}L := \{u \in \Sigma^* \mid \exists v \in K, vu \in L\}$
- $LK^{-1} := \{u \in \Sigma^* \mid \exists v \in K, uv \in L\}$

Lemma

1. $\widetilde{LK^{-1}} = (\tilde{K})^{-1}\tilde{L}$
2. *Si* $K \neq \emptyset$ *alors* $L \subseteq K^{-1}(K \cdot L)$
3. *L'inclusion inverse n'est pas vraie.*

1. $u \in \widetilde{LK^{-1}}$ ssi $\exists v \in \Sigma^*$ tq 1) $u = \tilde{v}$ et 2) $v \in LK^{-1}$. Or (2) ssi $\exists x \in K$ tq $vx \in L$ i.e. $\tilde{x}\tilde{v} \in \tilde{L}$. Donc (2) ssi $\exists y \in \tilde{K}$ tq $y\tilde{v} \in \tilde{L}$. Donc $u \in \widetilde{LK^{-1}}$ ssi $u \in (\tilde{K})^{-1}\tilde{L}$.
2. $K^{-1}(K \cdot L) := \{u \in \Sigma^* \mid \exists v \in K, vu \in KL\}$. Soit $u \in L$. Soit $v \in K$, alors $vu \in KL$, donc $u \in K^{-1}(K \cdot L)$
3. Soit $K := \{a, aa\}$ et $L := \{aa\}$. On a $aaaa \in KL$ et $a \in K$, donc $aaa \in K^{-1}(K \cdot L)$, mais $aaa \notin L$.

Opérations sur les langages (III)

Pour tout $L, K \subseteq \Sigma^*$ on pose :

- $K^{-1}L := \{u \in \Sigma^* \mid \exists v \in K, vu \in L\}$
- $LK^{-1} := \{u \in \Sigma^* \mid \exists v \in K, uv \in L\}$

Lemma

1. *(Rappel) Si $K \neq \emptyset$ alors $L \subseteq K^{-1}(K \cdot L)$*
2. *L et $K \cdot (K^{-1}L)$ sont incomparables, même si $K \neq \emptyset$.*

Proof.

- Soit $K := \{a, aa\}$ et $L := \{aa\}$. Alors $a \in K^{-1}L$ car $aa \in L$ et $a \in K$, donc $aaa \in K \cdot (K^{-1}L)$, car $aa \in K$. Cependant $aaa \notin L$.
- Soit $K := \{a\}$ et $L := \{b\}$. Alors $K^{-1}L = \emptyset$, donc $K \cdot (K^{-1}L) = \emptyset$, alors que $L \neq \emptyset$.

$\square$

Pour tout $u \in \Sigma^*$ on écrit $u^{-1}L$ au lieu de $\{u\}^{-1}L$, qu'on appelle résiduel de L par u.