IA: Inférence grammaticale - EISTI - ING 2

Yannick Le Nir

Ecole Internationale des Sciences du Traitement de l'Information

yannick.lenir@eisti.fr

Introduction

Théorie de l'apprentissage

Langages k-réversibles

Algorithmes d'apprentissage

Méthode des k suffixes

Introduction

Reconstruire un grammaire à partir d'exemples de phrases correctes.

- Grammaires linguistiques (TAL)
- Grammaires des langages de programmation (Reverse ingeniering)
- Grammaires de nucléides (Bio-informatique)
- Grammaires d'images (Reconnaissance de formes)

Introduction

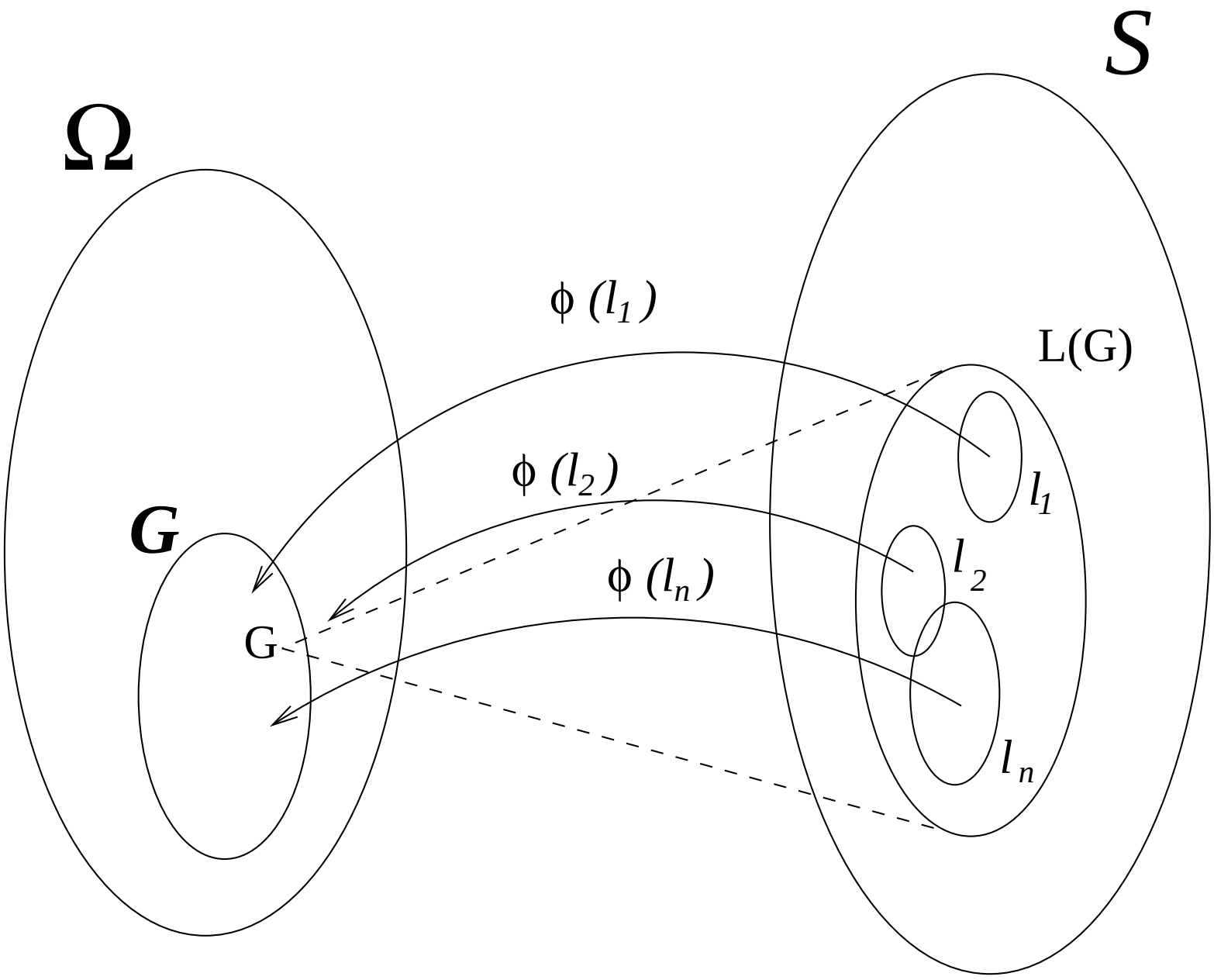
Théorie de l'apprentissage

Langages k-réversibles

Algorithmes d'apprentissage

Méthode des k suffixes

Théorie de l'apprentissage



Théorie de l'apprentissage

Définition

Soit $\langle \Omega, \mathcal{S}, L \rangle$ un système grammatical. Une *fonction d'apprentissage* est une fonction partielle qui associe aux ensembles finis non vides d'exemples positifs une grammaire,

$$\varphi : \bigcup_{k \geq 1} \mathcal{S}^k \rightarrow \Omega$$

Théorie de l'apprentissage

Définition

Soit $\langle \Omega, \mathcal{S}, L \rangle$ un système grammatical, $\langle s_i \rangle_{i \in \mathbb{N}} = \langle s_0, s_1, s_2, \cdots \rangle$ une séquence infinie d'exemples positifs appartenant à $\mathcal{S}$ et φ une fonction d'apprentissage définissant une grammaire $G_i = \varphi(\langle s_0, \cdots, s_i \rangle)$ pour tout $i \in \mathbb{N}$ tel que φ soit défini sur $\langle s_0, \cdots, s_i \rangle$.[1] La fonction φ *converge* vers G sur $\langle s_i \rangle_{i \in \mathbb{N}}$ si il existe $n \in \mathbb{N}$, tel que pour tout $i \geq n$, G_i est défini et $G_i = G$.

[1] $\langle s_0, \cdots, s_i \rangle$ désigne une séquence finie non vide. Si $i = 0$, on la remplace par $\langle s_0 \rangle$

Théorie de l'apprentissage

Définition

Soit $\langle \Omega, \mathcal{S}, L \rangle$ un système grammatical et $\mathcal{G} \subseteq \Omega$ une classe de grammaires. La fonction d'apprentissage φ *apprend* $\mathcal{G}$ si

- pour tout langage $L \in L(\mathcal{G}) = \{L(G) \mid \ \in \mathcal{G}\}$,
- et pour toute séquence infinie $\langle s_i \rangle_{i \in \mathbb{N}}$ énumérant L (i.e, $\{s_i \mid i \in \mathbb{N}\} = L$)

il existe $G \in \mathcal{G}$ tel que $L(G) = L$ et φ converge vers G sur $\langle s_i \rangle_{i \in \mathbb{N}}$.

Théorie de l'apprentissage

Définition

La classe de grammaire $\mathcal{G}$ est *apprenable* s'il existe une fonction d'apprentissage calculable qui apprend $\mathcal{G}$.

Théorème

Les grammaires régulières (type 3) ou algébrique (type 2) ne sont pas apprenables.

Langages k-réversibles

Définition

Un automate d'états finis admet un automate inverse, que l'on obtient en inversant les transitions entre les états de l'automate, et en remplaçant les états finaux en états initiaux (et vice -versa).

Introduction

Théorie de l'apprentissage

Langages k-réversibles

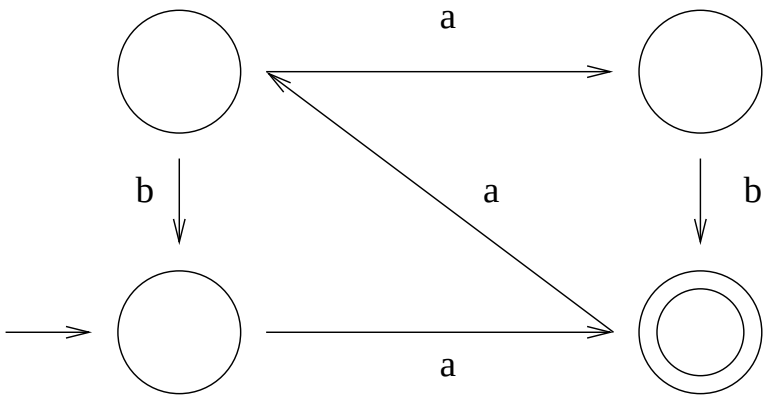
Algorithmes d'apprentissage

Méthode des k suffixes

Langages k-réversibles

Exemple d'automate inverse

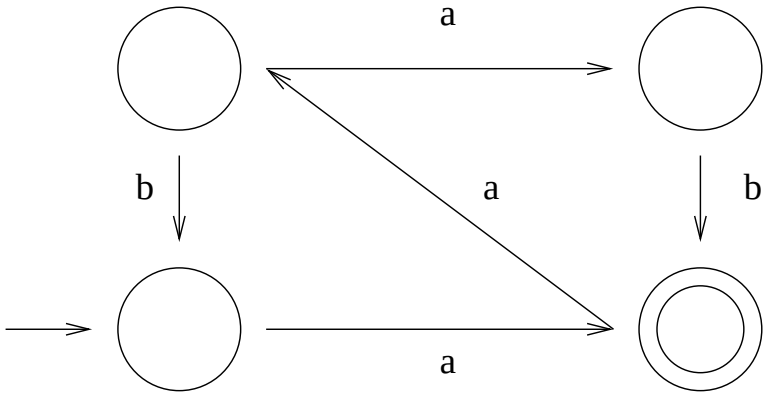
Soit l'automate déterministe suivant :



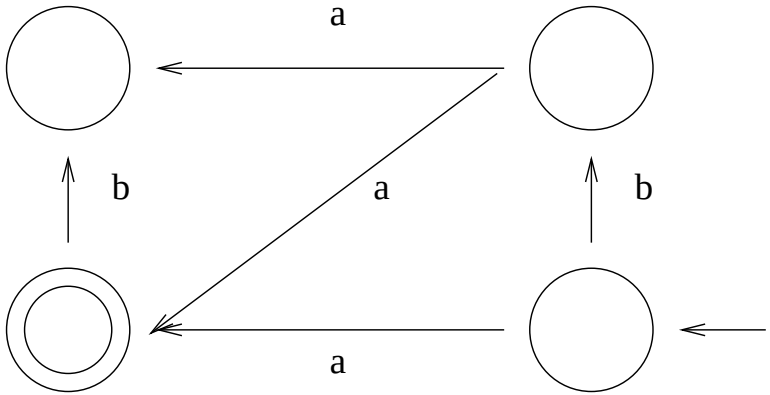
Langages k-réversibles

Exemple d'automate inverse

Soit l'automate déterministe suivant :



Son automate inverse est le suivant :



Langages k-réversibles

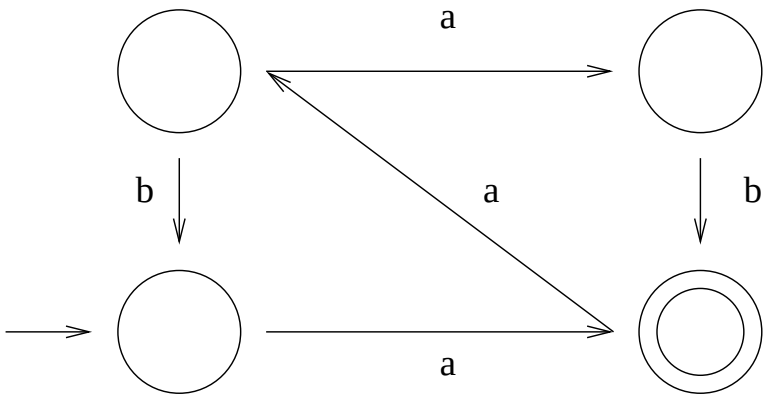
Définition

Un automate d'états finis est k-réversible ssi il est déterministe et l'automate inverse est également déterministe avec anticipation k :
$\forall q, q' \in Q, q \neq q', (q, q' \in Q_0) \vee (q, q' \in \delta(q, a)) \Rightarrow \neg\exists u \in \Sigma^k : (\delta(q, u) \neq \emptyset) \wedge (\delta(q', u) \neq \emptyset))$.

Langages k-réversibles

Exemple d'automate reversible

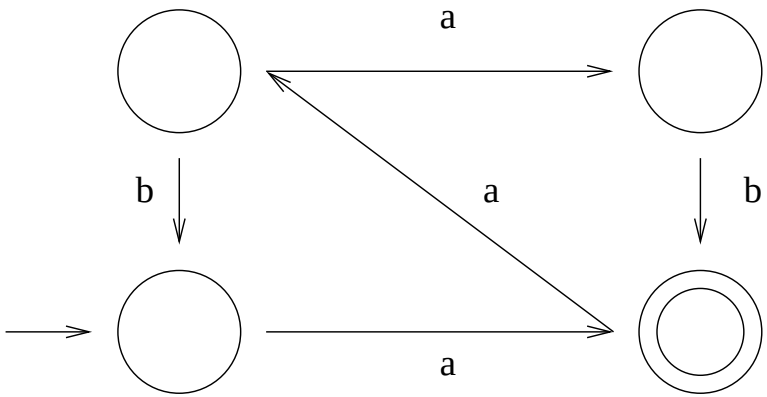
Reprenont l'automate précédant :



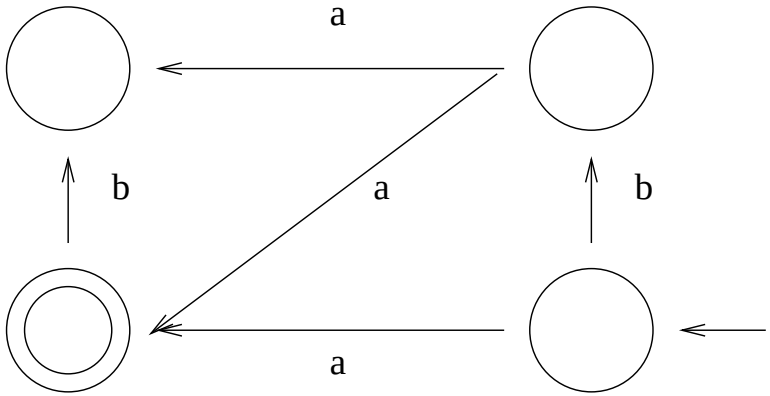
Langages k-réversibles

Exemple d'automate reversible

Reprenont l'automate précédant :



Son automate inverse est déterministe avec anticipation 1 :



Langages k-réversibles

Définition

Une grammaire régulière G est k-réversible si l'automate la reconnaissant est réversible.

Théorème

Les grammaire k-réversibles sont apprenables

Méthode d'agrégation des suffixes

Suffixe

L'ensemble des suffixes depuis un état de l'automate est l'ensemble des chaîne entre cet état et l'état final.

Distance

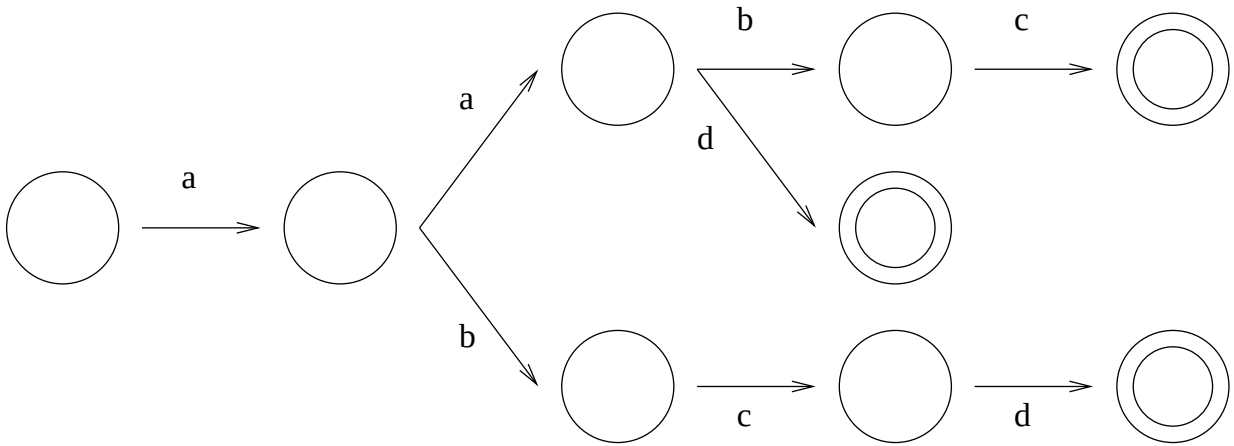
Distance d'inclusion entre deux ensembles de suffixes A et B :

- si $A \cap B \neq \emptyset$, $d(A, B) = min(|A|, |B|) - |A \cap B|$
- si $A \cap B = \emptyset$, $d(A, B) = \infty$

Méthode d'agrégation des suffixes

Exemple

A partir des exemples (aabc aad abcd), on obtient l'automate canonique suivant :



Méthode d'agrégation des suffixes

Algorithme

1. A partir de l'ensemble d'exemple, on construit l'automate canonique minimal.
2. Construction du tableau des distances entre ensemble de suffixes pour tous les états pris 2 à 2
3. Réduction du tableau en fusionnant les états ayant des distances finies
4. Répétition de l'étape précédente tant qu'il reste des distances infinies

Méthode des k suffixes

Algorithme

Soit $z \in T$ une chaîne telle que $zw \in R$ pour $w \in T$.
L'ensemble des k-suffixes de z par rapport à R constitue l'ensemble des sous-chaînes de longueur $\leq k$ ayant z comme préfixe dans R : $h(z, R, k) = \{w \mid zw \in R \text{ et } |w| \leq k\}$
Soit $Q = \{q \mid h(z, R, k) \text{ pour } z \in T^*\}$ et pour chaque $a \in T$, $\delta(q, a) = \{q' \in Q \mid q' = h(za, R, k) \text{ où } q = h(z, R, k)\}$
L'automate construit à partir de R a pour états des sous-ensembles de l'ensemble de toutes les k-suffixes pouvant être construits à partir de R.