Etude de la complexité moyenne en états des opérations rationnelles sur les langages finis

Frédérique Bassino, Laura Giambruno and Cyril Nicaud

LIPN, Université Paris-Nord
Dipartimento di Matematica e Informatica, Università di Palermo, Palermo
Institut Gaspard-Monge, Université Paris-Est

22 Mars 2011

Langages réguliers

- Soit L un langage régulier sur un alphabet fini.
- On a différentes façons de représenter un langage régulier : par automate déterministe, par automate non-déterministe ou par expression régulière
- Pour représenter de manière unique un langage on utilise son automate minimal
- On définit la complexité en états du langage L le nombre d'états de son automate minimal

Langages réguliers

- Soit L un langage régulier sur un alphabet fini.
- On a différentes façons de représenter un langage régulier : par automate déterministe, par automate non-déterministe ou par expression régulière
- Pour représenter de manière unique un langage on utilise son automate minimal
- On définit la complexité en états du langage L le nombre d'états de son automate minimal

Langages réguliers

- Soit L un langage régulier sur un alphabet fini.
- On a différentes façons de représenter un langage régulier : par automate déterministe, par automate non-déterministe ou par expression régulière
- Pour représenter de manière unique un langage on utilise son automate minimal
- On définit la complexité en états du langage L le nombre d'états de son automate minimal

Langages réguliers : notre étude en moyenne

- C’est intéressant de voir comment évolue cette ”mesure” sur les langages quand on applique les opérations rationnelles.
- Plusieurs résultats sont connus pour la complexité en états dans le pire cas pour les opérations rationnelles sur les langages représentés par automate.
- Dans ce travail on a considèrè les langages finis.

Notre travail

On fait l’étude en moyenne de la complexité en états des opérations rationnelles sur les langages finis, en considérant la distribution uniforme sur les ensembles avec un nombre fixé m de mots et de somme de longueurs n, avec n que tend vers l’infini.

Langages réguliers : notre étude en moyenne

- C'est intéressant de voir comment évolue cette "mesure" sur les langages quand on applique les opérations rationnelles.
- Plusieurs résultats sont connus pour la complexité en états dans le pire cas pour les opérations rationnelles sur les langages représentés par automate.
- Dans ce travail on a considèrè les langages finis.

Notre travail

On fait l'étude en moyenne de la complexité en états des opérations rationnelles sur les langages finis, en considérant la distribution uniforme sur les ensembles avec un nombre fixé m de mots et de somme de longueurs n, avec n que tend vers l'infini.

Langages réguliers : notre étude en moyenne

- C'est intéressant de voir comment évolue cette ”mesure” sur les langages quand on applique les opérations rationnelles.
- Plusieurs résultats sont connus pour la complexité en états dans le pire cas pour les opérations rationnelles sur les langages représentés par automate.
- Dans ce travail on a considèrè les langages finis.

Notre travail

On fait l'étude en moyenne de la complexité en états des opérations rationnelles sur les langages finis, en considérant la distribution uniforme sur les ensembles avec un nombre fixé m de mots et de somme de longueurs n, avec n que tend vers l'infini.

Problème

Soit X un ensemble fini de mots non vides et soit w un mot. Comment déterminer si $w \in X^*$?

- m est le nombre des mots dans $X = \{u_1, \cdots, u_m\}$
- n est la somme des longueurs des mots de X : $n = \sum_{i=1}^{m} |u_i|$
- $n \longrightarrow$ *longueur* de X

Problème

Soit X un ensemble fini de mots non vides et soit w un mot. Comment déterminer si $w \in X^*$?

- m est le nombre des mots dans $X = \{u_1, \cdots, u_m\}$
- n est la somme des longueurs des mots de X : $n = \sum_{i=1}^{m} |u_i|$
- $n \longrightarrow$ *longueur* de X

Problème

Soit X un ensemble fini de mots non vides et soit w un mot. Comment déterminer si $w \in X^*$?

- m est le nombre des mots dans $X = \{u_1, \cdots, u_m\}$
- n est la somme des longueurs des mots de X : $n = \sum_{i=1}^{m} |u_i|$
- $n \longrightarrow$ *longueur* de X

Constructions classiques

- On peut construire un automate non-déterministe $\mathcal{A}$ avec au plus n états reconnaissant X^*
- La complexité pour reconnaître un mot est en $\mathcal{O}(n \times |w|)$

- On déterminise $\mathcal{A} \Rightarrow \mathcal{B}$
- La complexité est $\mathcal{O}(|w|)$ une fois $\mathcal{B}$ calculé

Constructions classiques

- On peut construire un automate non-déterministe $\mathcal{A}$ avec au plus n états reconnaissant X^*
- La complexité pour reconnaître un mot est en $\mathcal{O}(n \times |w|)$

- On déterminise $\mathcal{A} \Rightarrow \mathcal{B}$
- La complexité est $\mathcal{O}(|w|)$ une fois $\mathcal{B}$ calculé

Croissance exponentielle

- Ellul, Krawetz, Shallit and Wand, Regular expressions : new results and open problems, *J. Autom. Lang. Combin. 10*, 2005.

Théorème

Pour chaque $h \geq 3$, il existe un ensemble X_h ayant h éléments et de taille $\Theta(h^2)$ tel que la complexité en états de $X_h{}^*$ est en $\Theta(h2^h)$.

La *complexité en états* d'un langage rationnel est le nombre d'état de son automate minimal.

Une solution plus efficace

- Clément, Duval, Guaiana, Perrin and Rindone, Parsing with a finite dictionary, *TCS 340*, 2005.

Théorème

Déterminer si w est dans X^* ou pas peut être décidé en temps $\mathcal{O}(m|w| + n)$.

Leur construction est inspirée de l'algorithme de Aho-Corasick.

Une solution plus efficace

- Clément, Duval, Guaiana, Perrin and Rindone, Parsing with a finite dictionary, *TCS 340*, 2005.

Théorème

Déterminer si w est dans X^* ou pas peut être décidé en temps $\mathcal{O}(m|w| + n)$.

Leur construction est inspirée de l'algorithme de Aho-Corasick.

Sur la concatenation : complexité dans le pire de cas

- C. Campeanu, K. Culik, K. Salomaa, S. Yu, State complexity of basic operations on finite languages, *Vol. 2214 in Lect. Notes Comput. Sci.*, 2001.

Théorème

Soient $\mathcal{A}_1$, $\mathcal{A}_2$ deux automate deterministe reconaissant des langages finis X_1, X_2, avec n_1 et n_2 états respectivement, et soit t le nombre d'états finals de $\mathcal{A}_1$. Alors il existe un automate deterministe $\mathcal{A}$ avec $\mathcal{O}(n_1 n_2^{t-1} + n_2^t)$ états tel que $L(\mathcal{A}) = L(\mathcal{A}_1)L(\mathcal{A}_2)$.

Pour $|A| = 2$ et $n_1 + 1 \geq n_2 > 2$ la borne supérieure est $(n_1 - n_2 + 3)2^{n_2-2} - 1$ et est atteinte.

Nos résultats en moyenne

Soient n, m entiers, $n >> m$ et A alphabet.

On considère ensembles X de A^* de longueur n avec m mots non vides, uniformément distribuées.

Complexité en moyenne des opérations rationnelles

- En moyenne, la complexité en états de l'étoile d'un ensemble X est linéaire par rapport à la longueur de X (avec équivalence pour $|A| \geq 3$)
- En moyenne, la complexité en états de l'union de la concaténation de deux ensembles X_1, X_2 est linéaire par rapport à la somme de les longueurs de X_1 et X_2

Complexité en moyenne de l'étoile : automates

- Soit A alphabet et $X \subset A^*$ ensemble fini de mots.
- Soit $\mathrm{Pr}(X)$ l'ensemble des préfixes des mots de X
- On définit $\mathcal{T}_X = (A, \mathrm{Pr}(X), T_X, \{\varepsilon\}, X)$, avec $T_X = \{(u, a, ua) \mid u \in \mathrm{Pr}(X),\ a \in A,\ ua \in \mathrm{Pr}(X)\}$
- $\mathcal{T}_X$ est un automate déterministe reconnaissant X

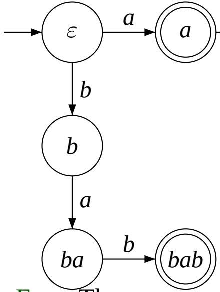


FIG.: The automaton $\mathcal{T}_X$ for $X = \{a, aba, bab\}$

Complexité en moyenne de l'étoile : automates

- On définit $\mathcal{A}_X = (A, \mathrm{Pr}(X), T_X \cup T, \{\varepsilon\}, X \cup \{\varepsilon\})$, avec $T = \{(u, a, a) \mid u \in X,\ a \in A \cap \mathrm{Pr}(X)\}$
- En particulier, pour chaque état final p, et pour chaque $a \in A$, si $\{a\}$ est dans $\mathcal{T}_X$, on ajoute à $\mathcal{T}_X$ un arc de p vers $\{a\}$ avec étiquette a.
- $\mathcal{A}_X$ est un automate non-déterministe reconnaissant X^*

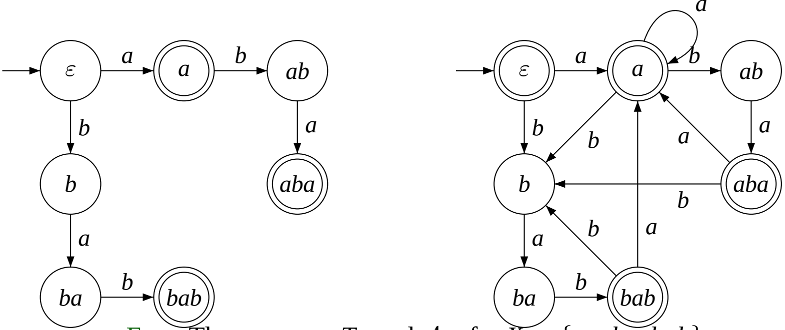
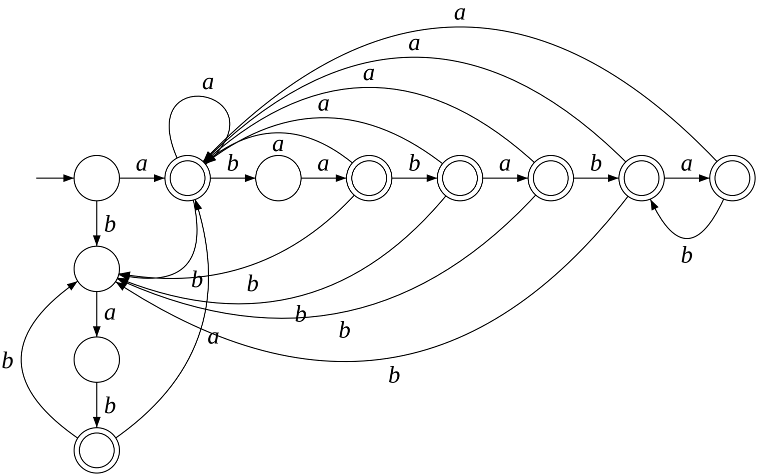


FIG.: The automata T_X and $\mathcal{A}_X$, for $X = \{a, aba, bab\}$

$\mathcal{D}_X$

- On déterminise l'automate précédent $\Rightarrow \mathcal{D}_X$



Propriétés de $\mathcal{D}_X$

- les états de l'automate non déterministe sont étiquetés avec les préfixes des mots de X
- par construction, si P est un état de $\mathcal{D}_X$, il est étiqueté avec un ensemble de préfixes de mots de X
- $P \cdot a$ peut seulement contenir a et des mots de la forme wa pour $w \in P$

Lemme

Une étiquette P d'un état de $\mathcal{D}_X$ est un ensemble de préfixes des mots de X tel que si w et w' sont dans P, alors soit w est un suffixe de w' soit w' est un suffixe de w.

Propriétés de $\mathcal{D}_X$

- les états de l'automate non déterministe sont étiquetés avec les préfixes des mots de X
- par construction, si P est un état de $\mathcal{D}_X$, il est étiqueté avec un ensemble de préfixes de mots de X
- $P \cdot a$ peut seulement contenir a et des mots de la forme wa pour $w \in P$

Lemme

Une étiquette P d'un état de $\mathcal{D}_X$ est un ensemble de préfixes des mots de X tel que si w et w' sont dans P, alors soit w est un suffixe de w' soit w' est un suffixe de w.

Propriétés de $\mathcal{D}_X$

- les états de l'automate non déterministe sont étiquetés avec les préfixes des mots de X
- par construction, si P est un état de $\mathcal{D}_X$, il est étiqueté avec un ensemble de préfixes de mots de X
- $P \cdot a$ peut seulement contenir a et des mots de la forme wa pour $w \in P$

Lemme

Une étiquette P d'un état de $\mathcal{D}_X$ est un ensemble de préfixes des mots de X tel que si w et w' sont dans P, alors soit w est un suffixe de w' soit w' est un suffixe de w.

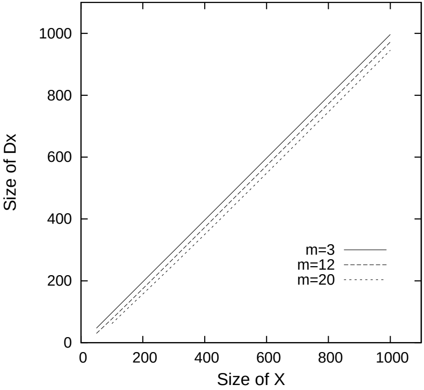
Propriétés de $\mathcal{D}_X$

- les états de l'automate non déterministe sont étiquetés avec les préfixes des mots de X
- par construction, si P est un état de $\mathcal{D}_X$, il est étiqueté avec un ensemble de préfixes de mots de X
- $P \cdot a$ peut seulement contenir a et des mots de la forme wa pour $w \in P$

Lemme

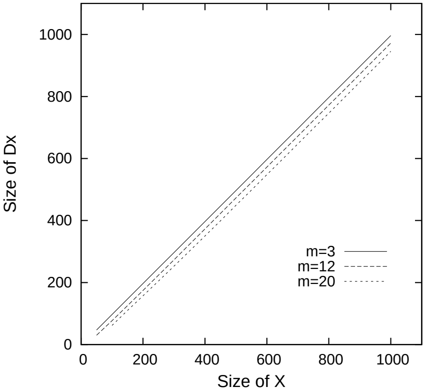
Une étiquette P d'un état de $\mathcal{D}_X$ est un ensemble de préfixes des mots de X tel que si w et w' sont dans P, alors soit w est un suffixe de w' soit w' est un suffixe de w.

Analyse de complexité moyenne : expériences



Dans ces expériences, le nombre d'états de D_X en moyenne semble être linéaire.

Analyse de complexité moyenne : expériences



Dans ces expériences, le nombre d'états de D_X en moyenne semble être linéaire.

Distribution sur les ensembles de mots

Différentes distributions utilisées :

- m fixé. Distribution uniforme sur les ensembles de m mots non vides dont la somme des longueurs est n.
- Distribution uniforme sur les ensembles de mots non vides dont la somme des longueurs est n. Trop de petits mots, X^* est souvent cofini.
- m fixé. Distribution uniforme sur ensembles de m mots non vides dont chaque mot est de longueur au plus n. Trop de grandes mots, X est un code préfixe avec très grande probabilité.

Distribution sur les ensembles de mots

Différentes distributions utilisées :

- m fixé. Distribution uniforme sur les ensembles de m mots non vides dont la somme des longueurs est n.
- Distribution uniforme sur les ensembles de mots non vides dont la somme des longueurs est n. Trop de petits mots, X^* est souvent cofini.
- m fixé. Distribution uniforme sur ensembles de m mots non vides dont chaque mot est de longueur au plus n. Trop de grandes mots, X est un code préfixe avec très grande probabilité.

Distribution sur les ensembles de mots

Différentes distributions utilisées :

- m fixé. Distribution uniforme sur les ensembles de m mots non vides dont la somme des longueurs est n.
- Distribution uniforme sur les ensembles de mots non vides dont la somme des longueurs est n. Trop de petits mots, X^* est souvent cofini.
- m fixé. Distribution uniforme sur ensembles de m mots non vides dont chaque mot est de longueur au plus n. Trop de grandes mots, X est un code préfixe avec très grande probabilité.

Distribution sur les ensembles de mots

Différentes distributions utilisées :

- m fixé. Distribution uniforme sur les ensembles de m mots non vides dont la somme des longueurs est n.
- Distribution uniforme sur les ensembles de mots non vides dont la somme des longueurs est n. Trop de petits mots, X^* est souvent cofini.
- m fixé. Distribution uniforme sur ensembles de m mots non vides dont chaque mot est de longueur au plus n. Trop de grandes mots, X est un code préfixe avec très grande probabilité.

Distribution sur les ensembles de mots

Différentes distributions utilisées :

- m fixé. Distribution uniforme sur les ensembles de m mots non vides dont la somme des longueurs est n.
- Distribution uniforme sur les ensembles de mots non vides dont la somme des longueurs est n. Trop de petits mots, X^* est souvent cofini.
- m fixé. Distribution uniforme sur ensembles de m mots non vides dont chaque mot est de longueur au plus n. Trop de grandes mots, X est un code préfixe avec très grande probabilité.

Distribution sur les ensembles de mots

Pour un fixé $m \geq 2$.

- Soit $\mathcal{S}et_{n,m}$ l'ensemble des ensembles de m mots non vides de somme de longueurs n
- On considère la distribution uniforme sur chaque n
- Le valeur moyenne de f est

$$\frac{1}{|\mathcal{S}et_{n,m}|} \sum_{X \in \mathcal{S}et_{n,m}} f(X^*)$$

- Asymptotique pour n quand n tend vers l'infini

Sur ensembles des mots

Une *séquence* de m mots est un m-uple $(u_1, \cdots, u_m)$ de mots.

- On peut travailler sur séquences $\mathcal{S}^{\neq}_{n,m}$ de m mots distincts, parce que il y a exactement $m!$ séquences correspondant à un ensemble donné

$$\frac{1}{|\mathcal{S}^{\neq}_{n,m}|} \sum_{X \in \mathcal{S}^{\neq}_{n,m}} f(X^*) = \frac{1}{m!\,|\mathcal{S}et_{n,m}|} \sum_{X \in \mathcal{S}et_{n,m}} m!\, f(X^*)$$

- On peut travailler sur les séquences $\mathcal{S}_{n,m}$ de m mots, parce que

$$|\mathcal{S}_{n,m}| = |\mathcal{S}^{\neq}_{n,m}| \left(1 + \mathcal{O}\left(\frac{1}{n^2}\right)\right)$$

Alors

$$\frac{1}{|\mathcal{S}et_{n,m}|} \sum_{X \in \mathcal{S}et_{n,m}} f(X^*) \approx \frac{1}{|\mathcal{S}_{n,m}|} \sum_{X \in \mathcal{S}_{n,m}} f(X^*)$$

Sur ensembles des mots

Une *séquence* de m mots est un m-uple $(u_1, \cdots, u_m)$ de mots.

- On peut travailler sur séquences $\mathcal{S}^{\neq}_{n,m}$ de m mots distincts, parce que il y a exactement $m!$ séquences correspondant à un ensemble donné

$$\frac{1}{|\mathcal{S}^{\neq}_{n,m}|} \sum_{X \in \mathcal{S}^{\neq}_{n,m}} f(X^*) = \frac{1}{m!\,|\mathcal{S}et_{n,m}|} \sum_{X \in \mathcal{S}et_{n,m}} m!\, f(X^*)$$

- On peut travailler sur les séquences $\mathcal{S}_{n,m}$ de m mots, parce que

$$|\mathcal{S}_{n,m}| = |\mathcal{S}^{\neq}_{n,m}| \left(1 + \mathcal{O}\left(\frac{1}{n^2}\right)\right)$$

  Alors

$$\frac{1}{|\mathcal{S}et_{n,m}|} \sum_{X \in \mathcal{S}et_{n,m}} f(X^*) \approx \frac{1}{|\mathcal{S}_{n,m}|} \sum_{X \in \mathcal{S}_{n,m}} f(X^*)$$

Complexité en moyenne de l'étoile : borne supérieure

Théorème

Pour la distribution uniforme sur les ensembles X de m mots non vides de somme de longueurs n, le nombre moyen d'états de $\mathcal{D}_X$ est en $\mathcal{O}(n)$.

Corollaire

La complexité en moyenne de X^* est linéaire.

Complexité en moyenne de l'étoile : borne supérieure

Théorème

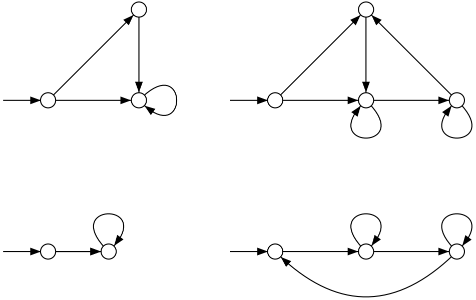
Pour la distribution uniforme sur les ensembles X de m mots non vides de somme de longueurs n, le nombre moyen d'états de $\mathcal{D}_X$ est en $\mathcal{O}(n)$.

Corollaire

La complexité en moyenne de X^* est linéaire.

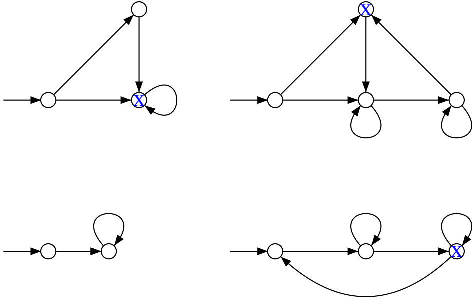
Idée de la preuve

$$\sum_{X \in \mathcal{S}_{n,m}} \text{taille de } \mathcal{D}_X = \sum_{P \text{ étiquette d'état}} \text{nombre de } X \text{ t.q. } P \text{ est un état de } \mathcal{D}_X$$



Idée de la preuve

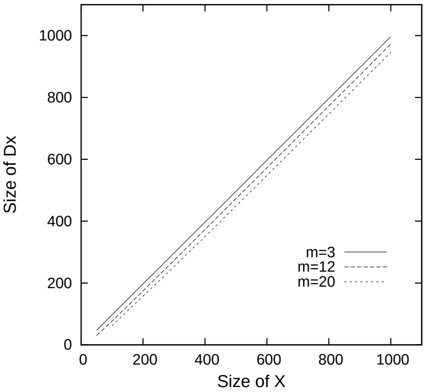
$$\sum_{X \in \mathcal{S}_{n,m}} \text{taille de } \mathcal{D}_X = \sum_{P \text{ étiquette d'état}} \text{nombre de } X \text{ t.q. } P \text{ est un état de } \mathcal{D}_X$$



Idée de la preuve

- Les étiquettes qui peuvent apparaître sont de la forme $\{x_1, \cdots, x_k\}$, avec x_i suffixe propre de x_{i+1}.
- Combinatoire sur les mots.
- Comptage de séquences avec contraintes.

A nouveau les expériences



Complexité en moyenne de l'étoile : équivalence à n

Théorème

Supposons que l'alphabet a au moins trois lettres. Pour la distribution uniforme sur ensembles X de m mots non vides de somme de longueurs n, la complexité en états de X^* est inferieure ou égal à $n + \mathcal{O}(1)$.

- Etiquettes qui pouvent apparaître sont de la forme $\{x_1, \cdots, x_k\}$, avec x_i suffixe propre de x_{i+1}. De plus, pour chaque x_i, x_j, il existe $w \in \mathit{Suff}(X)X^* \cap X^+$ tel que $x_i = wx_j$
- Combinatoire sur les mots.
- Comptage de séquences avec contraintes.

Complexité en moyenne de l'étoile : équivalence à n

Théorème

Supposons que l'alphabet a au moins trois lettres. Pour la distribution uniforme sur ensembles X de m mots non vides de somme de longueurs n, la complexité en états de X^* est inferieure ou égal à $n + \mathcal{O}(1)$.

- Etiquettes qui pouvent apparaître sont de la forme $\{x_1, \cdots, x_k\}$, avec x_i suffixe propre de x_{i+1}. De plus, pour chaque x_i, x_j, il existe $w \in \mathit{Suff}(X)X^* \cap X^+$ tel que $x_i = wx_j$
- Combinatoire sur les mots.
- Comptage de séquences avec contraintes.

Complexité en moyenne de la concaténation : résultat principal

Théorème

Pour la distribution uniforme sur les couples d'ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états moyenne de $X_1 \cdot X_2$ est inferieure ou égal à $(n_1 + n_2) + \mathcal{O}(1)$, quand n_1 et n_2 tendent vers infini.

Construction

- On définit $\mathcal{T}_{X_1}$ et $\mathcal{T}_{X_2}$ comme avant
- On définit $\mathcal{A}_{X_1 \cdot X_2}$ à partir de $\mathcal{T}_{X_1}$ et $\mathcal{T}_{X_2}$: pour chaque état final p de $\mathcal{T}_{X_1}$, pour chaque lettre a telle que $\{a\} \in Q_{X_2}$, on ajoute un arc de p vers $\{a\}$ avec étiquette a.
- $\mathcal{A}_{X_1 \cdot X_2}$ est un automate non-déterministe qui reconnaît $X_1 \cdot X_2$
- En déterminisant $\mathcal{A}_{X_1 \cdot X_2} \longrightarrow \mathcal{D}_{X_1 \cdot X_2}$

Construction : $\mathcal{A}_{X_1 \cdot X_2}$

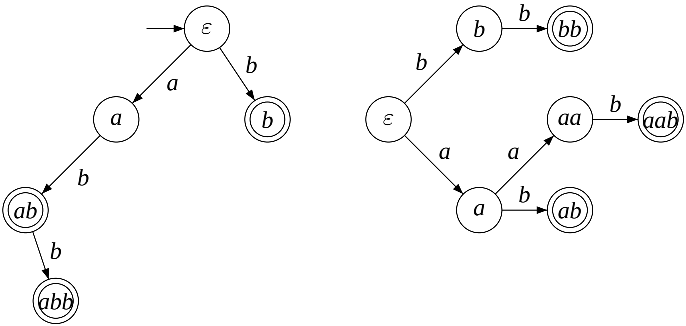


FIG.: Les automate $\mathcal{T}_{X_1}$ et $\mathcal{T}_{X_2}$, pour $X_1 = \{ab, abb, b\}$ et $X_2 = \{bb, ab, aab\}$.

Construction : $\mathcal{A}_{X_1 \cdot X_2}$

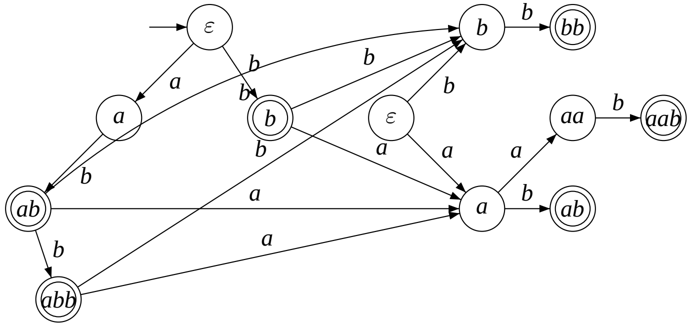


FIG.: L'automate $\mathcal{A}_{X_1 \cdot X_2}$, pour $X_1 = \{ab, abb, b\}$ et $X_2 = \{bb, ab, aab\}$.

Idée de la preuve

Lemme

Les états de $\mathcal{D}_{X_1 \cdot X_2}$ sont des couples (u, Z) de $(\mathrm{Pr}(X_1) \cup \emptyset) \times \mathcal{P}(\mathrm{Pr}(X_2))$ tels que

- Si $u \in \mathrm{Pr}(X_1)$, il existe un unique $Z \in \mathcal{P}(\mathrm{Pr}(X_2))$ tel que (u, Z) est un état de $\mathcal{D}_{X_1 \cdot X_2}$.
- Si $u = \emptyset$ et $Z = \{v_1, \cdots, v_\ell\}$, alors, si v est la longueur du plus long mot de Z, pour chaque $i \in \{1, \ldots, \ell\}$, $v = w_i v_i$

- L'idée de la preuve est similaire au cas de l'étoile.
- Comptage de séquences avec contraintes.

Complexité en moyenne de l’union : résultat principal

Théorème

Pour la distribution uniforme sur les couples d’ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états moyenne de $X_1 \cup X_2$ est inferieure ou égal à $(n_1 + n_2) + \mathcal{O}(1)$, quand n_1 et n_2 tendent vers l’infini.

ça vient de $|X_1 \cup X_2| \leq |X_1| + |X_2|$ et $\|X_1 \cup X_2\| \leq \|X_1\| + \|X_2\|$

Complexité en moyenne de l'union : résultat principal

Théorème

Pour la distribution uniforme sur les couples d'ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états moyenne de $X_1 \cup X_2$ est inferieure ou égal à $(n_1 + n_2) + \mathcal{O}(1)$, quand n_1 et n_2 tendent vers l'infini.

ça vient de $|X_1 \cup X_2| \leq |X_1| + |X_2|$ et $\|X_1 \cup X_2\| \leq \|X_1\| + \|X_2\|$

Bornes inferieures

Théorème

Pour la distribution uniforme sur ensembles X de $\mathcal{S}et_{n,m}$, la complexité en états de X^* est au moins $n + o(1)$.

Théorème

Pour la distribution uniforme sur les couples d'ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états en moyenne de $X_1 \cup X_2$ et de $X_1 \cdot X_2$ est bornée inférieurement par des fonctions équivalentes à $n_1 + n_2$, quand n_1 et n_2 tendent vers l'infini.

Bornes inferieures

Théorème

Pour la distribution uniforme sur ensembles X de $\mathcal{S}et_{n,m}$, la complexité en états de X^* est au moins $n + o(1)$.

Théorème

Pour la distribution uniforme sur les couples d'ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états en moyenne de $X_1 \cup X_2$ et de $X_1 \cdot X_2$ est bornée inférieurement par des fonctions équivalentes à $n_1 + n_2$, quand n_1 et n_2 tendent vers l'infini.

Bornes inferieurs : idée des preuves

- Il existe une famille de séquences $\mathcal{F} \subset S^{\neq}_{n,m}$ avec cardinalité asymptotiquement près de $S_{n,m}$
- La complexité en état d'un ensemble associée à une séquence dans $\mathcal{F}$ est asymptotiquement équivalente à n.

Bornes inferieurs : idée des preuves

- Il existe une famille de séquences $\mathcal{F} \subset S^{\neq}_{n,m}$ avec cardinalité asymptotiquement près de $S_{n,m}$
- La complexité en état d'un ensemble associée à une séquence dans $\mathcal{F}$ est asymptotiquement équivalente à n.

Résultats principaux

Théorème

Pour la distribution uniforme sur les ensembles X de $\mathcal{S}et_{n,m}$, la complexité en états moyenne de X^* est en $\Theta(n)$.

Théorème

Pour la distribution uniforme sur les couples d'ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états moyenne de $X_1 \cup X_2$ et de $X_1 \cdot X_2$ est en $\Theta(n_1 + n_2)$, quand n_1 et n_2 tendent vers l'infini.

Résultats principaux

Théorème

Pour la distribution uniforme sur les ensembles X de $\mathcal{S}et_{n,m}$, la complexité en états moyenne de X^* est en $\Theta(n)$.

Théorème

Pour la distribution uniforme sur les couples d'ensembles (X_1, X_2) de $\mathcal{S}et_{n_1,m_1} \times \mathcal{S}et_{n_2,m_2}$, la complexité en états moyenne de $X_1 \cup X_2$ et de $X_1 \cdot X_2$ est en $\Theta(n_1 + n_2)$, quand n_1 et n_2 tendent vers l'infini.

Complexité en temps

Theoreme

Pour $|A| \geq 3$, la complexité moyenne en temps pour la construction de$\mathcal{D}_X$ reconnaissant l'étoile d'un langage X en $\mathcal{S}et_{n,m}$ est en $\mathcal{O}(n \log n)$.

Théorème

Pour $|A| \geq 2$, la complexité moyenne en temps pour la construction de $\mathcal{D}_{X_1X_2}$ reconnaissant la concaténation des deux langages finis $X_1 \in \mathcal{S}et_{n_1,m_1}$ et $X_2 \in \mathcal{S}et_{n_2,m_2}$ est en $\mathcal{O}((n_1 + n_2) \log n_2)$.

Complexité en temps

Theoreme

Pour $|A| \geq 3$, la complexité moyenne en temps pour la construction de$\mathcal{D}_X$ reconnaissant l'étoile d'un langage X en $\mathcal{S}et_{n,m}$ est en $\mathcal{O}(n \log n)$.

Théorème

Pour $|A| \geq 2$, la complexité moyenne en temps pour la construction de $\mathcal{D}_{X_1X_2}$ reconnaissant la concaténation des deux langages finis $X_1 \in \mathcal{S}et_{n_1,m_1}$ et $X_2 \in \mathcal{S}et_{n_2,m_2}$ est en $\mathcal{O}((n_1 + n_2) \log n_2)$.

Idée de la preuve

- Comment calculer $Q = P \cdot a$?
- Est-ce que Q est déjà dans l'automate ?

Preuve :

- Arbres balancés pour chaque taille d'étiquettes
- Des types de sommes comme dans les calculs précédents.

Idée de la preuve

- Comment calculer $Q = P \cdot a$?
- Est-ce que Q est déjà dans l'automate ?

Preuve :

- Arbres balancés pour chaque taille d'étiquettes
- Des types de sommes comme dans les calculs précédents.

Conclusion et problems ouverts sur les langages finis

- En moyenne, les constructions de bases sont efficaces
- La complexité en états de X^* est équivalent à n en moyenne, pour un alphabet avec au moins 3 lettres.
- Est-ce vrai pour un alphabet binaire ? Oui, si $m = 2$
- Utiliser les chaines de Markov pour engendrer les mots ?

Conclusion et problems ouverts sur les langages finis

- En moyenne, les constructions de bases sont efficaces
- La complexité en états de X^* est équivalent à n en moyenne, pour un alphabet avec au moins 3 lettres.
- Est-ce vrai pour un alphabet binaire ? Oui, si $m = 2$
- Utiliser les chaines de Markov pour engendrer les mots ?

Conclusion et problems ouverts sur les langages finis

- En moyenne, les constructions de bases sont efficaces
- La complexité en états de X^* est équivalent à n en moyenne, pour un alphabet avec au moins 3 lettres.
- Est-ce vrai pour un alphabet binaire ? Oui, si $m = 2$
- Utiliser les chaines de Markov pour engendrer les mots ?

Conclusion et problems ouverts sur les langages finis

- En moyenne, les constructions de bases sont efficaces
- La complexité en états de X^* est équivalent à n en moyenne, pour un alphabet avec au moins 3 lettres.
- Est-ce vrai pour un alphabet binaire ? Oui, si $m = 2$
- Utiliser les chaines de Markov pour engendrer les mots ?

Complexité en états des opérations rationnelles sur les langages cofinis

- Nous avons étudié le pire cas de la complexité en états des opérations rationnelles sur les langages cofinis (langages tels que le langage complément est fini).
- Nous avons fourni des algorithmes pour calculer efficacement l'automate minimal.

Sur les langages cofinis

Un langage X es dit cofini si le langage complément $\overline{X} = A^* \setminus X$ est un langage fini.

- On défini la taille d'un langage cofini comme la taille de son ensemble complément.
- On note par $\|\overline{X}\|$ la somme des longueurs des mots de X : $\|\overline{X}\| = \sum_{i=1}^{m} |u_i|$

Sur les langages cofini

- Les langages cofinis sont stables pour les opérations rationnelles, notamment l'union, la concaténation et l'étoile.
- L'union d'un langage cofini et d'un langage régulier est un langage cofini.
- La concaténation d'un langage régulier X et d'un langage cofini Y est un langage cofini si et seulement si X contient un code préfixe maximal.

Nos résultats en détail : étoile et union

Théorème

Soient X_1 et X_2 des langages cofinis avec $\|\overline{X_1}\| = n_1$ et $\|\overline{X_2}\| = n_2$. L'union $X_1 \cup X_2$ est un langage cofini de complexité en états au plus $min(n_1, n_2) + 2$. L'automate minimal de $X_1 \cup X_2$ peut être calculé en temps $\mathcal{O}(n_1 + n_2)$.

Théorème

Pour tout langage cofini X avec $\|\overline{X}\| = n$, la complexité en états dans le pire cas de X^* est au plus $n + 2$. Il existe un algorithme pour calculer l'automate minimal de X^* en temps quadratique.

Nos résultats en détail : concaténation

Soit $\mathcal{S}et_{n,m}$ l'ensemble des ensembles de m mots non-vides dont la somme de longueurs est n.

Théorème

Soient X_1 et X_2 langages cofini tels que $X_1 \in Set_{n_1,m_1}$ et $X_2 \in Set_{n_2,m_2}$. La complexité en états de $X_1 \cdot X_2$ est au plus $n_1 + 1 + min(2m_2, n_2 + 2)$. Si $\varepsilon \in X_2$ (resp. $\varepsilon \in X_1$), la complexité en états de $X_1 \cdot X_2$ es au plus $n_1 + 2$ (resp. $n_2 + 2$). L'automate minimal de $X_1 \cdot X_2$ peut être calculé en temps $\mathcal{O}(n_1 + n_2)$.

Conclusion et problèmes ouverts sur les langages cofinis

- Les opérations rationnelles appliquées aux langages cofinis produisent des langages à faible complexité en états, et leur automate minimal peut être calculé rapidement.
- Concaténation d'un langage régulier et d'un langage cofini ?