Automates et mots infinis

Thibaut Verron

Décembre 2009

1 Mots infinis

1.1 Rappels : mots et langages

Définition 1 (Mot)**.** Soit A un ensemble fini, dénombrable ou indénombrable, appelé alphabet. Un mot sur A est une suite finie d'éléments de A ou lettres, notée $a_0 a_1 \cdots a_n$. Il existe un mot vide, noté ε. On note A^* l'ensemble des mots sur A, A^+ l'ensemble des mots non vides sur A.

L'ensemble A^* est muni de l'opération de concaténation, associative et admettant comme élément neutre ε. Ceci munit A^* d'une structure de monoïde.

Définition 2 (Morphisme)**.** Un morphisme de mots est une application $A^* \to B^*$ préservant la structure de monoïde.

Définition 3 (Opérations sur les langages)**.** Un langage L est un sous-ensemble de A^*. L'ensemble des langages est muni en particulier des opérations rationnelles :
- réunion $X + Y = X \cup Y$
- produit $XY = \{xy \mid x \in X \text{ et } y \in Y\}$
- étoile $X^* = \{x_1 \cdots x_n \mid n \geq 0 \text{ et } \forall i, x_i \in X\}$

On pose aussi $L^+ = LL^*$.

Définition 4 (Langage rationnel)**.** L'ensemble des langages rationnels est le plus petit ensemble contenant le langage vide, les lettres, et stable par les trois opérations rationnelles (réunion, produit, étoile).

1.2 Mots infinis

Définition 5 (Mot infini)**.** Un mot infini sur A est une suite infinie de lettres, notée $a_0 a_1 \cdots a_n \cdots$. On note A^ω l'ensemble des mots infinis sur A et on pose $A^\infty = A^* \cup A^\omega$.

La concaténation de mots finis s'étend en une opération $A^* \times A^\infty \to A^\infty$, de sorte qu'on peut définir le produit d'un sous-ensemble de A^* par un sous-ensemble de A^∞. On peut également définir des morphismes de mots infinis. L'opération de réunion s'étend sans problème aux langages de A^∞.

On définit une nouvelle opération appelée itération infinie $A^* \to A^\omega$ par

$$X^\omega = \{x_0 x_1 \cdots \mid \forall i \in \mathbb{N}, x_i \in X \setminus \{\varepsilon\}\}$$

Exemple. Soit $u = a_0 a_1 \cdots a_n$, $\{u\}^\omega = a_0 a_1 \cdots a_n a_0 a_1 \cdots a_n \cdots$.

Proposition 1. *Pour tous* $X, Y \subset A^*$ *:*
- $(X+Y)^\omega = (X^*Y)^\omega + (X+Y)^* X^\omega$
- $(XY)^\omega = X(YX)^\omega$
- $\forall n > 0, (X^n)^\omega = (X^+)^\omega = X^\omega$
- $XX^\omega = X^+ X^\omega = X^\omega$

Définition 6 (Langages ω-rationnels)**.** La classe des langages ω-rationnels de A^∞ est le plus petit ensemble $\mathcal{R}$ de parties de A^∞ tel que :
- $\emptyset \in \mathcal{R}$ et $\forall a \in A, \{a\} \in \mathcal{R}$
- $\mathcal{R}$ est stable par union finie
- $\mathcal{R}$ est stable par produit :

$$\text{si } X \subset A^*, Y \subset A^\infty, X \in \mathcal{R} \text{ et } Y \in \mathcal{R} \Rightarrow XY \in \mathcal{R}$$

- $\mathcal{R}$ est stable par étoile et itération infinie

Il est noté $\mathcal{R}at(A^\omega)$.

2 Reconnaissance de langages ω-rationnels

2.1 Automates de Büchi

Les automates de Büchi représentent le plus simple des moyens pour reconnaître des ensembles de mots infinis.

Définition 7. Un automate de Büchi est un quintuplet $\mathcal{A} = (Q, A, E, I, F)$, parfois noté $\mathcal{A} = (E, I, F)$ où :
- Q est un ensemble (supposé fini dans la suite) d'états
- A est l'alphabet
- E est un ensemble de transitions
- I et F sont des sous-ensembles de Q appelés ensembles des états initiaux et finaux

Dans la suite, on notera pour $p, q \in Q$ et $w \in A^*$, $p \xrightarrow{w} q$ s'il existe un chemin de p à q étiqueté w, et $p \xrightarrow[\mathrm{F}]{w} q$ s'il existe un chemin de p à q étiqueté w et visitant au moins un état final.

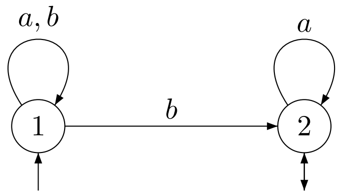

FIGURE 1 – Exemple d'automate non déterministe

Définition 8. L'ensemble des mots reconnus par un automate de Büchi est l'ensemble des mots définissant un chemin initial (partant d'un état initial) et final (visitant F une infinité de fois). Cet ensemble est noté $L^\omega(\mathcal{A})$.

On rappelle que $L^*(\mathcal{A})$ est l'ensemble des mots finis reconnus par l'automate, et on pose $L^+(\mathcal{A}) = L^*(\mathcal{A}) \setminus \{\varepsilon\}$.

Exemple. Soit $\mathcal{A}$ l'automate représenté en figure 1.
On a $L^+(\mathcal{A}) = (a+b)^+ = A^* \setminus \{\varepsilon\}$ et $L^\omega(\mathcal{A}) = (a+b)^* a^\omega$, ensemble des mots comportant un nombre fini de b. Cet automate n'est pas déterministe.

Théorème 2 (Kleene). *Un sous-ensemble de A^∞ est reconnaissable si et seulement si il est ω-rationnel.*

Démonstration. La preuve est analogue à celle du théorème de Kleene pour les mots finis. □

2.2 Clôture des langages reconnaissables

Proposition 3 (Stabilité par morphisme et morphisme inverse)**.**
Soit $\varphi : A^\omega \to B^\omega$ un morphisme de mots.
- *Si X est un sous-ensemble reconnaissable de A^ω, $\varphi(X)$ est un sous-ensemble reconnaissable de B^ω.*
- *Si X est un sous-ensemble reconnaissable de B^ω, $\varphi^{-1}(X)$ est un sous-ensemble reconnaissable de A^ω.*

Démonstration. La stabilité par morphisme découle du fait qu'un morphisme est caractérisé par l'image de ses lettre, comme dans le cas fini.

Pour l'image réciproque, à partir d'un automate de Büchi reconnaissant X, on peut facilement construire un automate reconnaissant $\varphi^{-1}(X)$. Soit $\mathcal{B} = (Q, A, E, I, F)$ un automate de Büchi reconnaissant X. Posons

$$\mathcal{A} = (Q \times \{0,1\}, A, E', I \times \{0\}, F \times \{0\})$$

avec $E' = E_1 \cup E_2$ et

$$E_1 = \Big\{ \big((q,i), a, (q',0) \big) \mid \text{il existe un chemin dans } \mathcal{B} \text{ de } q \text{ à } q' \\ \text{étiqueté } \varphi(a), \text{ ne visitant aucun état de } F \Big\}$$

et

$$E_2 = \left\{ \big((q,i), a, (q',1) \big) \mid p \xrightarrow[\mathrm{F}]{\varphi(a)} q \right\}$$

Cet automate reconnaît bien $\varphi^{-1}(X)$. □

Théorème 4 (Büchi)**.** *L'ensemble des langages ω-rationnels est clos par toutes les opérations booléennes (union, intersection et complémentation).*

La partie facile de ce théorème est donnée par l'énoncé suivant :

Proposition 5. *L'ensemble des langages ω-rationnels est clos par réunion et intersection.*

Démonstration. Stabilité par réunion : vraie par définition.

Stabilité par intersection : Soient L_1 et L_2 deux langages ω-rationnels, ils sont reconnus respectivement par les deux automates de Büchi $\mathcal{A}_1 = (Q_1, A, E_1, I_1, F_1)$ et $\mathcal{A}_2 = (Q_2, A, E_2, I_2, F_2)$.

$$I = I_1 \times I_2 \times \{1\}$$
$$F = Q_1 \times F_2 \times \{1\}$$
$$E = \Big\{ \Big((p_1, p_2, s), a, (q_1, q_2, t) \Big) \mid (p_1, a, q_1) \in E_1,\ (p_1, a, q_1) \in E_1 \text{ et } t = 0$$
$$\text{si et seulement si } (s = 1, p_2 \in F_2 \text{ et } q_1 \notin F_1) \text{ ou } (s = 0 \text{ et } q_1 \notin F_1) \Big\}$$

On peut vérifier que cet automate reconnaît $L_1 \cap L_2$: si on visite un état de F, cela signifie déjà qu'on serait en train de visiter un état de F_2 dans $\mathcal{A}_2$. A présent, supposons que l'on visite une infinité de fois F. Plaçons-nous à un de ces états $(p, q, 1)$, dans un chemin. Si entre $(p, q, 1)$ et l'état de F qui sera le prochain visité, on visite un état où $t = 0$, repasser à $t = 1$ implique qu'on visite un état de F_1. Sinon, cela signifie que l'état suivant immédiatement $(p, q, 1)$ sera de la forme $(p', q', 1)$, or $q \in F_2$ donc $p' \in F_1$. Ceci montre que tout mot reconnu par A est dans $L_1 \cap L_2$, la réciproque se montre en remontant le raisonnement.

□

La stabilité par complémentation requiert des définitions supplémentaires, ainsi que le théorème de Ramsey.

Définition 9 (Factorisation Ramseyenne)**.** Soit $u \in A^\infty$. Une factorisation de u est une suite $(u_n)_{n\in\mathbb{N}}$ de mots telle que $u = u_0u_1u_2\cdots$.

Soit $\varphi : A^+ \to E$. Une factorisation $u = xy_0y_1y_2\cdots$ est dite ramseyenne si et seulement si il existe $e \in E$ tel que pour tous i et j tels que $i \leq j$, $\varphi(y_iy_{i+1}\cdots y_j) = e$.

Théorème 6. *Soit $\varphi : A^+ \to E$. Tout mot infini de A^ω admet une factorisation ramseyenne pour φ.*

Démonstration. Cela se déduit du théorème de Ramsey par compacité. □

Proposition 7. *L'ensemble des langages ω-rationnels est stable par complémentation.*

Démonstration. Soit $\mathcal{A} = (Q, A, E, I, F)$ un automate de Büchi acceptant un langage X. On considère la relation $\sim$ sur A^* définie par

$$w \sim w' \Leftrightarrow \begin{cases} \forall p, q \in Q, p \xrightarrow{w} q \Leftrightarrow p \xrightarrow{w'} q \\ \forall p, q \in Q, p \xrightarrow[\mathrm{F}]{w} q \Leftrightarrow p \xrightarrow[\mathrm{F}]{w'} q \end{cases}$$

C'est une congruence : une relation d'équivalence compatible avec la concaténation. $A^*/\sim$ est un ensemble fini, de cardinal au plus $3^{|Q|^2}$.

Lemme 1. *Les classes d'équivalences sont des langages rationnels.*

Démonstration. Une classe d'équivalence est l'image réciproque d'un singleton par la projection canonique (qui est un morphisme). $\square$

Lemme 2. *Soient K, L deux classes pour $\sim$. Alors $KL^\omega \cap X = \emptyset$ ou $KL^\omega \subset X$.*

Démonstration. Supposons $KL^\omega \cap X \neq \emptyset$, et soient

$$w = \underbrace{u_0}_{\in K} \cdot \underbrace{u_1}_{\in L} \cdot \underbrace{u_2}_{\in L} \cdots \in KL^\omega \cap X$$

et

$$w' = \underbrace{u'_0}_{\in K} \cdot \underbrace{u'_1}_{\in L} \cdot \underbrace{u'_2}_{\in L} \cdots \in KL^\omega$$

Alors il existe $i \in I$, $q_0, q_1 ... \in Q$ tels que, par exemple,

$$i \xrightarrow{u_0} q_0 \xrightarrow[\mathrm{F}]{u_1} q_1 \xrightarrow{u_2} q_2 \cdots$$

chemin visitant une infinité de fois l'ensemble des états finaux. Mais K et L étant des classes de congruence, les mêmes états sont visités par le chemin étiqueté w', en visitant autant d'états finaux. Donc $w' \in X$, soit $KL^\omega \subset X$. $\square$

Il reste à remarquer que l'ensemble des langages de la forme KL^ω recouvre A^* : cela découle immédiatement du théorème de Ramsey. En effet, soit $x \in A^*$, obtenir une factorisation ramseyenne de x pour la projection canonique sur $A^*/\sim$ donne directement la factorisation de x souhaitée. Donc $A^\omega \setminus X = \bigcup_{KL^\omega \cap X = \emptyset} KL^\omega$, union finie d'ensembles rationnels, donc $A^\omega \setminus X$ est rationnel. $\square$

2.3 Déterminisation

Définition 10. Un automate de Büchi est dit déterministe si ses transitions sont déterministes et si l'automate a exactement un état initial. Dans ce cas, chaque mot définit au plus un chemin initial.

Dans le cas des mots finis, la démonstration de la stabilité des langages rationnels par complémentation se fait facilement en déterminisant l'automate. Le fait est que pour les mots infinis, les automates de Büchi déterministes ont un pouvoir de reconnaissance strictement inférieur à celui des automates de Büchi quelconques. On peut en effet facilement caractériser les langages reconnaissables par un automate déterministe. Commençons par introduire un nouvel opérateur.

Définition 11. Pour tout $L \subset A^*$, on pose

$$\overrightarrow{L} = \{u \in A^\omega \mid u \text{ a une infinité de préfixes dans } L\}$$

Exemple.

1. Si L est fini, $\overrightarrow{L} = \emptyset$.
2. Si $L = a^*b$, $\overrightarrow{L} = \emptyset$.
3. Si $L = (a+b)^*b = (a*b)^+$ (ensemble des mots finissant par un b), $\overrightarrow{L} = (a^*b)^\omega$ (ensemlbe des mots contenant une infinité de b).

Théorème 8. *Soit X un sous-ensemble de A^ω. Les assertions suivantes sont équivalentes :*

1. *X est reconnaissable par un automate de Büchi déterministe.*
2. *Il existe $L \subset A^+$ tel que $X = \overrightarrow{L}$.*

Démonstration. ($1 \Longrightarrow 2$) On peut montrer que si $\mathcal{A} = (Q, A, E, i, F)$ est un automate de Büchi déterministe, $L^\omega(\mathcal{A}) = \overrightarrow{L^+(\mathcal{A})}$ En effet, si $u \in L^\omega(\mathcal{A})$, u définit le chemin $q_0, q_1, q_2 \cdots$, et comme u est reconnu, il existe $n_0 < n_1 < \cdots$ tels que pour tout k, $q_{n_k} \in F$. Alors les mots $u_k = a_0 a_1 \cdots a_{n_k - 1}$ sont dans $L^+(\mathcal{A})$ et sont préfixes de u, par conséquent $L^\omega(\mathcal{A}) \subset \overrightarrow{L^+(\mathcal{A})}$. L'inclusion réciproque est évidente, ce qui montre (1) $\Longrightarrow$ (2).

($2 \Longrightarrow 1$)Soit L un sous-ensemble de A^+ tel que $X = \overrightarrow{L}$, alors L est reconnu par l'automate déterministe $\mathcal{A} = (A^*, A, \cdot, \varepsilon, L)$, où la fonction de transition est la concaténation à droite :

$$\forall u \in A^*, \forall a \in A, u \cdot a = ua$$

Alors $\mathcal{A}$, vu comme un automate de Büchi, reconnaît bien X. $\square$

Avec les notations précédentes, L est reconnaissable si et seulement si X est reconnaissable par un automate déterministe fini (il est facile d'adapter la démonstration).

Cette caractérisation permet de démontrer qu'il existe des langages ω-rationnels non reconnaissables par un automate de Büchi déterministe.

En effet, considérons l'ensemble $X = (a+b)^* a^\omega$ de l'exemple précédent. Il n'existe pas de langage L tel que $X = \overrightarrow{L}$. En effet, sinon, puisque $ba^\omega \in X$, il existe n_1 tel que $ba^{n_1} \in L$. Comme $ba^{n_1}ba^\omega \in X$, il existe n_2 tel que $ba^{n_1}ba^{n_2} \in L$. Et ainsi de suite, on construit une suite $(n_k)_{k\in\mathbb{N}}$ tel que pour tout k, $ba^{n_1}\cdots ba^{n_k} \in L$. Et alors, le mot $ba^{n_1}ba^{n_2}\cdots$ a une infinité de préfixes dans L mais contient une infinité de b, donc n'appartient pas à X. Le langage X ne peut donc pas être reconnu par un automate de Büchi déterministe (même infini).

Il existe un résultat analogue au théorème de déterminisation pour les mots finis, mais il faut considérer un mode de reconnaissance plus puissant.

Définition 12 (Automates de Muller)**.** Un automate de Muller est un quintuplet $\mathcal{A} = (Q, A, E, i, \mathcal{T})$, parfois noté $\mathcal{A} = (E, i, \mathcal{T})$ où (Q, A, E) est un automate fini *déterministe*, i l'état initial et $\mathcal{T}$ un ensemble de parties de Q, appelé table de l'automate.

Etant donné un chemin p dans l'automate, on note $\mathrm{Inf}(p)$ l'ensemble des états visités une infinité dans le chemin. Un mot est reconnu par l'automate $\mathcal{A}$ si et seulement si il définit un chemin p partant de i et tel que $\mathrm{Inf}(p) \in \mathcal{T}$.

Théorème 9. *L'ensemble des mots infinis reconnus par un automate de Muller est ω-rationnel.*

En particulier, les automates de Muller reconnaissent au plus autant de langages que les automates de Buchi finis.

Démonstration. Soit $\mathcal{A} = (Q, A, E, i, \mathcal{T})$ un automate de Muller. On vérifie aisément que $L^\omega(\mathcal{A}) = \bigcup\limits_{T\in\mathcal{T}} L^\omega(E, i, \{T\})$, ce qui nous permet, puisque la classe des langages ω-rationnels est stable par union finie, de ne considérer que le cas où $\mathcal{T}$ est un singleton. Soit alors $\mathcal{T} = \{T\}$, où $T = \{t_0, t_1, \ldots, t_k\}$. On pose D la restriction de la fonction de transition à T, c'est-à-dire $D = E \cap T \times A \times T$.

Posons $X = L^+(E, i, \{t_0\})$, $Y_0 = L^+(D, \{t_1\}, \{t_0\})$, et ainsi de suite jusqu'à $Y_k = L^+(D, \{t_k\}, \{t_0\})$. X est l'ensemble des mots définissant un chemin dans $\mathcal{A}$ allant de i à t_0, Y_1 est l'ensemble des mots définissant un chemin de t_0 à t_1 ne passant que par des états de T, et ainsi de suite. On va montrer que $L^\omega(\mathcal{A}) = X(Y_0Y_1\cdots Y_K)^\omega$, ce qui prouvera immédiatement que $L^\omega(\mathcal{A})$ est ω-rationnel.

L'inclusion $X(Y_0Y_1\cdots Y_K)^\omega \subset L^\omega(\mathcal{A})$ est évidente. Soit u dans $L^\omega(\mathcal{A})$. Donc il existe $q_0, q_1, q_2, \cdots \in Q$, $a_0, a_1, a_2, \cdots \in A$ tels que $u = a_0a_1a_2\cdots$, $q_0 \xrightarrow{a_0} q_1, q_1 \xrightarrow{a_1} q_2, \ldots$ et Inf(p)=T. En particulier, il existe n_0 tel que $q_{n_0} = t_0$ et pour tout $n \geq n_0$, on a $q_n \in T$. En itérant cette recherche pour tous les éléments de T, on peut trouver une suite infinie $n_0 < n_1 < n_2 < \cdots$ telle que la suite $q_{n_0}, q_{n_1}, q_{n_2}, \ldots$ est égale à la suite $t_0, t_1 \ldots t_k, t_0, \ldots, t_k, t_0 \ldots$. En découpant le chemin en $n_0, n_1 \ldots$, on obtient la factorisation voulue.

□

Exemple. Considérons le langage $X = (a+b)^* a^\omega$ des mots comportant un nombre fini de b, précédemment étudié. Il est ω-rationnel et on connaît un automate de Büchi reconnaissant le langage X. On sait aussi qu'il n'existe pas d'automate de Büchi déterministe reconnaissant X. En revanche, il existe des automates de Muller reconnaissant X : un exemple en est l'automate représenté en figure 2, avec la table $\mathcal{T} = \{\{2\}\}$.

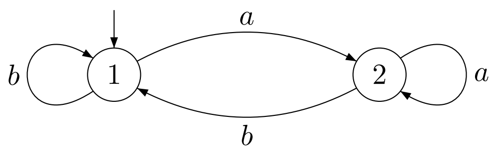

FIGURE 2 – Automate de Muller reconnaissant X

Ce résultat ce généralise par le théorème de McNaughton, dont l'énoncé est plus intéressant que la démonstration.

Théorème 10 (McNaughton)**.** *Toute partie reconnaissable de A^ω peut être reconnue par un automate de Muller.*

Références

[1] Dominique Perrin - Jean-Eric Pin, **Infinite words**, 2004