PUBLICATIONS DU DÉPARTEMENT DE MATHÉMATIQUES DE LYON

LONG VAN DO

Codes infinitaires et automates non-ambigus

Publications du Département de Mathématiques de Lyon, 1985, fascicule 2B
« Compte rendu des journées infinitistes », , p. 97-107
<http://www.numdam.org/item?id=PDML_1985___2B_97_0>

CODES INFINITAIRES ET AUTOMATES NON-AMBIGUS

par Do Long Van

On donne des caractérisations des codes infinitaires en termes d'automates non ambigus.

One gives some characterizations of infinitary codes in terms of non-ambiguous automata.

1. Rappel de quelques définitions et résultats nécessaires.

Etant donné un alphabet A on note comme d'habitude A^* l'ensemble des mots finis sur A, A^ω l'ensemble des mots infinis sur A, et $A^\infty = A^* \cup A^\omega$. On appelle langage <u>infinitaire</u> (resp. <u>finitaire</u>, <u>purement infinitaire</u>) toute partie X de A^∞ (resp. A^*, A^ω).

Un automate sur A,

$$\mathcal{A} = (Q \; , \; I \; , \; T_{fin} \; , \; T_{inf}),$$

est composé

- d'un ensemble Q <u>d'états</u>,
- d'un ensemble $I \subseteq Q$ d'<u>états initiaux</u>,
- d'un ensemble $T_{fin} \subseteq Q$ d'<u>états finaux</u>,
- d'un ensemble $T_{inf} \subseteq Q$ d'<u>états infinis</u>, et
- d'un ensemble $\mathcal{F}_{\mathcal{A}} = Q \times A \times Q$ de <u>flèches</u>.

Une flèche (p,a,q) est aussi noté $p \xrightarrow{a} q$. Un <u>chemin</u> dans $\mathcal{A}$ est une suite (finie ou infinie) $c = c_1 c_2 \ldots$ de flèches $c_i : p_i \xrightarrow{a_i} q_i$ consécutives (i.e. $q_i = p_{i+1}$, $i \geqslant 1$). Le mot $\alpha = a_1 a_2 \ldots$ est appelé <u>étiquette</u> du chemin c. Dans la suite on notera par $c : p \xrightarrow{f} q$ un chemin fini c d'étiquette u qui commence à l'état p et se termine à l'état q ; et par $c : p \overset{u}{\twoheadrightarrow} T_{inf}$ un chemin infini c d'étiquette u qui commence à p et passe infiniment souvent l'ensemble T_{inf}. On pose ensuite :

$$L^* (\mathcal{A}) = \{f \in A^* \mid \exists\, c : i \xrightarrow{f} t \ , \ i \in I \ , \ t \in T_{fin}\},$$

$$L^{\omega} (\mathcal{A}) = \{u \in A^{\omega} \mid \exists\, c : i \overset{u}{\twoheadrightarrow} T_{inf} \ , \ i \in I\} \ ,$$

$$L(\mathcal{A}) = L^*(\mathcal{A}) \cup L^{\omega} (\mathcal{A}) \ ,$$

et on dit que le langage $L(\mathcal{A})$ est <u>reconnu</u> par $\mathcal{A}$.

Un automate $\mathcal{A} = (Q , I, T_{fin}, T_{inf})$ est <u>émondé</u> s'il satisfait les conditions suivantes :

- tout $q \in Q$ est accessible de I ;
- tout $q \in Q$ est coaccessible de $T_{fin} \cup T_{inf}$;
- l'ensemble T_{inf} est inclus dans $\{q \in Q \mid \exists\, c : q \twoheadrightarrow T_{inf}\}$.

<u>PROPOSITION 1.1</u> (M. Nivat /2/). *Pour tout automate* $\mathcal{A} = (Q, I, T_{fin}, T_{inf})$ *il existe un automate émondé* $\mathcal{A}^e = (Q^e, I^e, T^e_{fin}, T^e_{inf})$ *avec* $Q^e \subseteq Q$, $I^e \subseteq I$, $T^e_{fin} \subseteq T_{fin}$, $T^e_{inf} \subseteq T_{inf}$, $\mathcal{F}_{\mathcal{A}^e} \subseteq \mathcal{F}_{\mathcal{A}}$ *tel que* $L(\mathcal{A}^e) = L(\mathcal{A})$.

L'automate $\mathcal{A}^e$ est appelé <u>partie émondée</u> de $\mathcal{A}$.

Un automate $\mathcal{A} = (Q, I, T_{fin}, T_{inf})$ est <u>fini</u> si les ensembles Q et A sont finis. Un langage X est dit <u>reconnaissable</u> s'il existe un automate fini qui reconnait X.

PROPOSITION 1.2. (A. Arnold /3/). *Tout langage purement infinitaire reconnaissable est reconnu par un automate fini* $\mathcal{A} = (Q, I, \emptyset, T_{inf})$ *non ambigu dans le sens suivant :*

$$\forall u \in A^{\omega} \ : \text{Card } \{c : i \xrightarrow{u}\!\!\!\twoheadrightarrow T_{inf} \mid i \in I\} \leqslant 1.$$

Soit A un alphabet. On munit A^{∞} d'un produit de la manière suivante :

$$\forall u \in A^{\omega} \ \forall \alpha \in A^{\infty} \ : u.\alpha = u,$$

$$\forall f \in A^{*} \ \forall \alpha \in A^{\infty} \ : f.\alpha = f\,\alpha\,,$$

où $f\,\alpha$ est le résultat de la concaténation des mots f et α. Alors A^{∞} est un monoïde ayant pour élément neutre le mot vide noté ε. Pour toute partie X de A^{∞} on note X^{*} le sous-monoïde de A^{∞} engendré par X et on pose $X^{+} = X^{*} - \varepsilon$. Dans la suite au lieu de $\alpha.\beta$ on écrira simplement $\alpha\beta$, $\alpha, \beta \in A^{\infty}$.

Pour toute partie X de A^{∞} on note $X_{fin} = X \cap A^{*}$, $X_{inf} = X \cap A^{\omega}$, et on pose

$$\begin{cases} X^{(1)} = X, \\ X^{(n)} = \{(x_1, \dots, x_n) \mid x_1, \dots, x_{n-1} \in X_{fin},\ x_n \in X\}\ ,\ n = 2 \end{cases} .$$

Une partie X de A^{∞} est un <u>code (infinitaire)</u> (voir /4/ - /8/) si pour tous $n, m \geqslant 1$ et pour tous $(x_1, \dots, x_n) \in X^{(n)}$, $(x'_1, \dots, x'_m) \in X^{(m)}$, l'égalité

$$x_1 \dots x_n = x'_1 \dots x'_m$$

implique $n = m$ et $x_i = x'_i$ $(i = 1, \dots, n)$. On appelle <u>quasi-libre</u> tout sous-monoide M de A^{∞} engendré par un code X. Le code X est appelé la <u>base</u> de M.

Pour toutes parties X et Y de A on note

$$X\,Y^{-1} = \{\alpha \in A^{\infty} \mid \exists\, \beta \in Y : \ \alpha\beta \in X\}\ ;$$

$$Y^{-1}X = \{\alpha \in A^{\infty} \mid \exists\, \beta \in Y : (\beta\,\alpha \in X)\ \&\ ((|\beta| = \omega) \Rightarrow (\alpha = \varepsilon))\}\ .$$

Un sous-monoïde M de A^{∞} est <u>libérable</u> si $M^{-1}M \cap MM^{-1} \subseteq M$.

Soit M un sous-monoïde de A^∞ et soient $u,v \in M_{inf}$. On dit que u est contenu dans v, noté $u < v$, s'il existe $f \in (M_{fin}-\varepsilon)$ tel que $u = f v$. On appelle chaine toute suite croissante $u_1 < u_2 < \dots$ d'éléments de M_{inf}. On dit qu'un sous-monoïde M satisfait à la condition de chaine finie si toute chaine d'éléments de M_{inf} est finie.

Un ensemble générateur d'un sous-monoïde M est minimum s'il est inclus dans tout ensemble générateur de M.

Une partie X de A^∞ est distinguée si $X_{inf} \cap X^+_{fin}X_{inf} = \emptyset$. Evidemment sont distingués tout langage finitaire ainsi que tout code infinitaire.

PROPOSITION 1.3 (cf. /7/). *Pour tout sous-monoïde M de A^∞, les conditions suivantes sont équivalentes :*

(i) M est quasi-libre ;

(ii) M est libérable et possède un ensemble générateur minimum distingué ;

(iii) M est libérable et satisfait à la condition de chaine finie.

PROPOSITION 1.4 (cf. /8/). *Soit M un sous-monoïde quasi-libre de A^∞ et X sa base. Alors*

$$X = (M - \varepsilon) - (M_{fin} - \varepsilon)(M - \varepsilon).$$

Donc X est reconnaissable si M l'est.

2. PROPOSITIONS AUXILIAIRES.

On appelle série formelle toute application $S : A^\infty \to \mathbb{N} \cup \{\omega\}$. En notant (S,α) l'image du mot α par S, on peut écrire la série S sous la forme :

$$S = \sum_{\alpha \in A^\infty} (S,\alpha)\alpha \quad .$$

Une série S est caractéristique si pour tout $\alpha \in A^\infty$, $(S,\alpha) \leqslant 1$.

Pour tout langage X on appelle <u>série caractéristique</u> de X la série $\underline{X}$ ainsi définie

$$(\underline{X},\alpha) = \begin{cases} 1 & \text{si } \alpha \in X \\ 0 & \text{si } \alpha \notin X \end{cases}$$

Soient S et T deux séries formelles. Alors S + T et S T sont définies par :

$$(S + T , \alpha) = (S , \alpha) + (T , \alpha) ;$$

$$(S\,T ,\alpha) = \sum_{\substack{\beta,\gamma \in A^{\infty} \\ \beta\gamma = \alpha \\ (|\beta|=\omega) \Rightarrow (\gamma=\varepsilon)}} (S , \beta)(T , \gamma) .$$

L'addition de séries formelles est évidemment commutative et associative. On vérifie sans peine que la multiplication est aussi associative. On peut donc parler des puissances S^n de la série S, où $S^0 = \underline{\{\varepsilon\}}$. Puis on définit $S^* = \sum_{n \geq 0} S^n$. La proposition suivante donne une caractérisation des codes infinitaires en termes de séries formelles :

<u>PROPOSITION 2.1.</u> *Pour toute partie* X *de* $A^{\infty} - \varepsilon$, X *est un code ssi* $(\underline{X})^*$ *est une série caractéristique.*

A chaque automate $\mathcal{A} = (Q, I, T_{fin}, T_{inf})$ on associe canoniquement un automate $\mathcal{A}^*$ défini par la construction en des étapes que voici :

1) Posons $\mathcal{B} = (Q \cup \omega, \omega, \omega, \emptyset)$ avec $\omega \notin Q$ et $\mathcal{F}_{\mathcal{B}} = \mathcal{F}_{\mathcal{A}} \cup \mathcal{I} \cup \mathcal{T} \cup \mathcal{O}$

où

$$\mathcal{I} = \{(\omega,a,q) \mid \exists\, i \in I : (i,a,q) \in \mathcal{F}_{\mathcal{A}} \},$$

$$\mathcal{T} = \{(p,a,\omega) \mid \exists\, t \in T_{fin} : (p,a,t) \in \mathcal{F}_{\mathcal{A}}\}$$

$$\mathcal{O} = \{(\omega,a,\omega) \mid \exists\, i \in I, \exists\, t \in T_{fin} : (i,a,t) \in \mathcal{F}_{\mathcal{A}}\}$$

2) Posons $\mathscr{C} = (Q \cup \delta, \delta, \emptyset, T_{inf})$ avec $\delta \notin Q$ et $\mathscr{F}_{\mathscr{C}} = \mathscr{F}_{\mathscr{A}} \cup \mathscr{K}$,
où

$\mathscr{K} = \{(\delta, a, q) \mid \exists\, i \in I : (i,a,q) \in \mathscr{F}_{\mathscr{A}}\}$.

3) Notons $\mathscr{D}$ l'union disjointe de $\mathscr{B}$ et $\mathscr{C}$ en collant ω avec δ.

4) L'automate $\mathscr{A}^*$ est par définition la partie émondée de $\mathscr{D}$.

On vérifie facilement que l'automate $\mathscr{A}^*$ est de la forme $\mathscr{A}^* = (Q,1,1,T_{inf})$ et satisfait à la condition suivante :

$$\forall\, (c : 1 \twoheadrightarrow T_{inf}) : 1 \notin \mathrm{Inf}(c) \qquad (*)$$

où Inf(c) est l'ensemble des états infiniment répétés sur le chemin c.

Maintenant à chaque automate $\mathscr{A} = (Q,I,T_{fin},T_{inf})$ on associe une série formelle $|\mathscr{A}|$ définie comme suit :

$$(|\mathscr{A}|,\alpha) = \begin{cases} \mathrm{Card}\ \{ c : i \xrightarrow{\alpha} t \mid i \in I,\ t \in T_{fin}\} & \text{si } \alpha \in A^* , \\ \mathrm{Card}\ \{ c : i \overset{\alpha}{\twoheadrightarrow} T_{inf} \mid i \in I \} & \text{si } \alpha \in A^{\omega} . \end{cases}$$

On a :

<u>PROPOSITION 2.2.</u> *Soit X une partie de $A^{\infty} - \varepsilon$.*

(i) Si $\mathscr{A}$ est un automate tel que $L(\mathscr{A}) = X$, alors
$$L(\mathscr{A}^*) = X^* ;$$

(ii) Si $\mathscr{A}$ est un automate tel que $|\mathscr{A}| = \underline{X}$, alors
$$|\mathscr{A}^*| = (\underline{X})^* .$$

Un automate $\mathscr{A} = (Q,I,T_{fin},T_{inf})$ est dit <u>non-ambigu</u> si

$$\forall\, p,q \in Q\ \ \forall f \in A^* : \mathrm{Card}\ \{ c : p \xrightarrow{f} q\} \leqslant 1 ,$$
$$\forall\, p \in Q\ \ \forall\, u \in A^{\omega} : \mathrm{Card}\ \{ c : p \overset{u}{\twoheadrightarrow} T_{inf}\} \leqslant 1.$$

PROPOSITION 2.3. *Soit* $\mathcal{A} = (Q,i,t,T_{inf})$ *un automate émondé ayant un unique état initial* i *et un unique état final* t. *Alors* $\mathcal{A}$ *est non-ambigu ssi* $|\mathcal{A}|$ *est une série caractéristique.*

Des propositions 2.1, 2.2, et 2.3 il résulte la

PROPOSITION 2.4. *Soit* X *une partie de* $A^{\infty} - \varepsilon$ *et soit* $\mathcal{A}$ *un automate tel que* $|\mathcal{A}| = \underline{X}$. *Alors* X *est un code ssi* $\mathcal{A}^*$ *est non-ambigu.*

En utilisant la proposition 1.3 on peut démontrer la

PROPOSITION 2.5. *Soit* $\mathcal{A} = (Q,1,1,T_{inf})$ *un automate non-ambigu ayant un seul état final confondu avec un seul état initial et vérifiant la condition (*). Alors* $L(\mathcal{A})$ *est un sous-monoïde quasi-libre de* A^{∞} .

En vertu de la proposition 1.2 on a :

PROPOSITION 2.6. *Tout langage reconnaissable* X *est reconnu par un automate tel que* $|\mathcal{A}| = \underline{X}$.

3. Résultats principaux.

Comme une conséquence immédiate de la proposition 2.1 (i) on a le résultat suivant qui caractérise les sous-monoïdes reconnaissables :

THEOREME 3.1. *Soit* X *une partie de* A^{∞}. *Alors* X *est un sous-monoïde reconnaissable ssi il existe un automate fini* $\mathcal{A} = (Q,1,1,T_{inf})$ *ayant un seul état final confondu avec un seul état initial et vérifiant la condition (*) qui reconnait* X.

Le résultat suivant dont la démonstration sera donnée dans la section 4 caractérise les codes reconnaissables en termes d'automates non-ambigus.

THEOREME 3.2. *Soit* X *une partie de* $A^{\infty} - \varepsilon$. *Les conditions suivantes sont équivalentes :*

(i) X *est un code reconnaissable ;*

(ii) X *est reconnaissable et* $\forall \mathscr{A}(|\mathscr{A}| = \underline{X} \Rightarrow \mathscr{A}$ *est non-ambigu)*

(iii) X *est l'ensemble générateur minimum de* X* *et il existe un automate fini non-ambigu* $\mathscr{A} = (Q,1,1,T_{inf})$ *vérifiant la condition* (*) *qui reconnait* X*.

(iv) X *est l'ensemble générateur minimum distingué de* X* *et il existe un automate fini non-ambigu* $\mathscr{A} = (Q,1,1,T_{inf})$ *qui reconnait* X* ,

Comme une conséquence du théorème 3.2 et de la proposition 2.5 on a

THEOREME 3.3. *Soit* X *une partie de* A^{∞}. *Alors* X *est un sous-monoïde quasi-libre reconnaissable ssi il existe un automate fini non-ambigu* $\mathscr{A} = (Q,1,1,T_{inf})$ *ayant un seul état final confondu avec un seul état initial et vérifiant la condition* (*) *qui reconnait* X.

A chaque partie X de $A^* - \varepsilon$ on associe canoniquement un automate $\mathscr{A}_D(X) = (Q, I, T_{fin}, \emptyset)$ défini par

$$Q = \{(f,g) \in A^* \times A^* \mid fg \in X\} ;$$

$$I = \{\varepsilon\} \times X \; ; \; T_{fin} = X \times \{\varepsilon\} \; ;$$

$$\mathscr{F}_{\mathscr{A}_D(X)} = \{(f,ag) \xrightarrow{a} (fa,g) \mid fag \in X, \; a \in A \} \quad .$$

Soit maintenant X une partie de $A^{\infty} - \varepsilon$ telle que

$$X_{inf} = \bigcup_{j \in J} Y_j Z_j^{\omega} \; ; \; Y_j, Z_j \subseteq A^* ,$$

où $Z_j^{\omega} = \{w \in A^{\omega} \mid w = g_1 g_2 \ldots, g_i \in Z_j\}$. On associe à X un automate $\mathscr{A}_D^*(X)$, appelé <u>automate en pétales</u> de X, défini par la construction en des étapes suivantes :

1) Soient $\mathscr{A}_D(Y_j - \varepsilon) = (Q_j, I_j, T_j, \emptyset)$ et $[\mathscr{A}_D(Z_j)]^* = (Q'_j, r_j, r_j, \emptyset)$ avec $Q_j \cap Q'_j - \emptyset$.

Posons

$$\mathcal{A}_D(Y_j,Z_j) = (Q_j \cup Q'_j, I'_j, \emptyset, r_j),$$

où $$I'_j = \begin{cases} I_j & \text{si } \varepsilon \notin Y_j \\ I_j \cup \{r_j\} & \text{si } \varepsilon \in Y_j \end{cases},$$

et

$$\mathcal{F}_{\mathcal{A}_D}(Y_j,Z_j) = \mathcal{F}_{\mathcal{A}_D}(Y_j-\varepsilon) \cup \mathcal{F}_{\mathcal{A}_D}(Z_j) \cup \mathcal{G}$$

avec

$$\mathcal{G} = \{(p,a,r_j) \mid \exists\, t \in T_j : (q,a,t) \in \mathcal{F}_{\mathcal{A}_D}(Y_j-\varepsilon)\} \quad .$$

2) Notons $\mathcal{A}_D(X)$ l'union disjointe de $\mathcal{A}_D(X_{fin})$ et des $\mathcal{A}_D(Y_j,Z_j)$:

$$\mathcal{A}_D(X) = \mathcal{A}_D(X_{fin}) \cup \bigcup_{j \in J} \mathcal{A}_D(Y_j,Z_j) \quad ;$$

3) L'automate $\mathcal{A}^*_D(X)$ est par définition $[\mathcal{A}_D(X)]^*$.

Une famille $(Y_j,Z_j)_{j \in J}$ de parties de A^* est <u>non-ambigue</u> si pour chaque mot α de A^ω il existe au plus une factorisation de la forme $\alpha = fg_1g_2 \dots$ avec $f \in Y_j$, $g_i \in Z_j$, $j \in J$.

Le résultat suivant dont la démonstration se fait d'une façon similaire à celle du théorème 3.2. caractérise une classe de codes infinitaires en termes d'automates non ambigus.

<u>THEOREME 3.4.</u> *Soit X une partie de $A^\infty - \varepsilon$ telle que*

$$X_{inf} = \bigcup_{j \in J} Y_j Z_j^\omega$$

où la famille $(Y_j,Z_j)_{j \in J}$ est non-ambigue. Alors les conditions suivantes sont équivalentes :

(i) X est un code ;

(ii) $\forall \mathscr{A} \; (|\mathscr{A}| = \underline{X} \Rightarrow \mathscr{A}^*$ *est non-ambigu);*

(iii) L'automate en pétales $\mathscr{A}_D^*(X)$ *est non-ambigu ;*

(iv) X *est l'ensemble générateur minimum de* X* *et il existe un automate non-ambigu* $\mathscr{A} = (Q,1,1,T_{inf})$ *vérifiant la condition* (*) *qui reconnait* X*;

(v) X *est l'ensemble générateur minimum distingué de* X* *et il existe un automate non-ambigu* $\mathscr{A} = (Q,1,1,T_{inf})$ *reconnaissant* X*.

4. Démonstration du théorème 3.2.

La démonstration se fait d'après le schéma que voici :

(i) $\Leftrightarrow$ (ii) $\Rightarrow$ (i)+(ii) $\Rightarrow$ (iii) $\Rightarrow$ (iv) $\Rightarrow$ (i).

(i) $\Leftrightarrow$ (ii) : l'implication (i) $\Rightarrow$ (ii) est par la proposition 2.4 ; l'implication inverse (ii) $\Rightarrow$ (i) est par les propositions 2.4 et 2.6.

(i) + (ii) $\Rightarrow$ (iii) : d'après la proposition 2.6 il existe un automate $\mathscr{B}$ tel que $|\mathscr{B}| = \underline{X}$. Posons $\mathscr{A} = \mathscr{B}^*$, alors $\mathscr{A}$ est de la forme $\mathscr{A} = (Q,1,1,T_{inf})$, vérifie la condition (*) et reconnait X*. Par (ii) $\mathscr{A}$ est non-ambigu ; par (i) X est l'ensemble générateur minimum de X*.

(iii) $\Rightarrow$ (iv) : il suffit de montrer que X est distingué. Par la proposition 2.5 X* est quasi-libre, donc son ensemble générateur minimum X est un code, et par conséquent distingué.

(iv) $\Rightarrow$ (i) : en utilisant la non-ambiguité de l'automate $\mathscr{A}$ il n'est pas difficile de démontrer que X* est libérable. Ayant X pour un ensemble générateur minimum distingué, X* est par la proposition 1.3 quasi-libre. Donc X est un code, et par la proposition 1.4. il est reconnaissable.

Ces résultats sont des extensions naturelles de quelques résultats dans /1/.

L'auteur exprime sa gratitude à Monsieur le Professeur D. PERRIN dont les conseils et les encouragements ont créés d'excellentes conditions de cette étude.

Références.

/1/. J. BERSTEL, D. PERRIN, M.P. SCHUTZENBERGER, Théorie des codes, Publication du LITP, N°33 (1982).

/2/. M. NIVAT, Behaviours of synchronized systems of processes, Publication du LITP, N°64 (1981).

/3/. A. ARNOLD, Rational ω-languages are non-ambiguous, Theoretical Computer Science, 26 (1983), 221-223.

/4/. L.V. DO, Codes avec des mots infinis, RAIRO Informatique théorique, 16 (1982), 4, 371-386.

/5/. L.V. DO, Sous-monoïdes et codes avec des mots infinis, Semigroup Forum, 26 (1983), 1, 75-87.

/6/. L.V. DO, Ensembles code-compatibles et une généralisation du théorème de Sardinas-Patterson, Publication du LITP, N°10 (1982), (à paraître dans "Theoretical Computer Science").

/7/. L.V. DO, Caractérisations combinatoires des sous-monoïdes engendrés par un code infinitaire, Preprint Series Hanoi (1984), N°6.

/8/. L.V. DO, Sur les ensembles générateurs minimums des sous-monoïdes de A^{∞}, Preprint Series Hanoi (1984), n°11.

Institut de Mathématiques
P/O Box 631 BO HO, Hanoi, Vietnam
et
LITP, Université de Paris VII,
2, place Jussieu
75221 PARIS Cedex 05 (France)