MPRI, Fondations mathématiques de la théorie des automates

Olivier Carton, Jean-Éric Pin

Examen du 2 mars 2015, 12h 45. Durée: 2h 30, notes de cours autorisées

⋆ ⋆ ⋆

Avertissement : On attachera une grande importance à la clarté, à la précision et à la concision de la rédaction.

Soit A un alphabet. Les langages considérés dans ce problème sont des langages réguliers de A^*.

On dit qu'un langage S *sépare* deux langages K et L si $K \subseteq S$ et $L \cap S = \emptyset$. Plus généralement, si $\mathcal{F}$ est une **algèbre de Boole de langages** de A^*, on dit que K et L sont $\mathcal{F}$*-séparables* s'il existe un langage de $\mathcal{F}$ qui sépare K et L.

On s'intéresse au *problème de la séparation* pour $\mathcal{F}$: peut-on décider si deux langages réguliers donnés sont $\mathcal{F}$-séparables?

1. Généralités et exemples

Question 1. Montrer que L et L^c sont $\mathcal{F}$-séparables si et seulement si $L \in \mathcal{F}$.

Question 2. Soient $(K_i)_{1 \leqslant i \leqslant n}$ et $(L_j)_{1 \leqslant j \leqslant m}$ deux familles finies de langages réguliers. Montrer que les langages $\bigcup_{1 \leqslant i \leqslant n} K_i$ et $\bigcup_{1 \leqslant i \leqslant m} L_j$ sont $\mathcal{F}$-séparables si et seulement si chaque paire (K_i, L_j) est $\mathcal{F}$-séparable.

Question 3. On prend pour $\mathcal{F}$ la classe des langages sans-étoile. Montrer que les langages $(a^2)^*$ et $(a^2)^*a$ ne sont pas $\mathcal{F}$-séparables, mais que les langages $(ab)^+$ et ba^* le sont.

Question 4. On prend pour $\mathcal{F}$ la classe des langages testables par morceaux. Montrer que les langages $(ab)^+$ et $(ba)^+$ ne sont pas $\mathcal{F}$-séparables, mais que les langages $(a^2)^*$ et $(b^2)^*b$ le sont. Montrer qu'il n'existe cependant pas de $\mathcal{F}$-séparateur minimal pour ces langages.

Dans les questions 5 et 6, on prend pour $\mathcal{F}$ la classe des langages dont le monoïde syntactique est idempotent et commutatif. On rappelle que ce sont les langages qui sont combinaisons booléennes de langages de la forme A^*aA^*, avec $a \in A$, ou encore les langages qui sont combinaisons booléennes de langages de la forme B^*, où $B \subseteq A$.

Question 5. Si $A = \{a, b, c\}$, montrer que $(ab)^+$ et $(ac)^+$ sont $\mathcal{F}$-séparables mais que $(a(b+c))^+$ et $(ba)^+$ ne sont pas $\mathcal{F}$-séparables.

Question 6. Montrer que le problème de la séparation pour $\mathcal{F}$ est décidable.

2. Ensembles ponctuels

Soit M un monoïde et soit $\mathbf{V}$ une variété de monoïdes finis. On dit qu'une partie X de M est $\mathbf{V}$*-ponctuelle* si, pour tout morphisme relationnel $\tau : M \to T$, avec $T \in \mathbf{V}$, il existe un élément $t \in T$ tel que $X \subseteq \tau^{-1}(t)$.

Question 7. Soit M le semigroupe $\{1, a, b, 0\}$ dans lequel le produit de deux éléments différents de 1 est égal à 0. On prend pour $\mathbf{V}$ la variété des semigroupes idempotents et commutatifs. Montrer que les parties $\{a, 0\}$ et $\{b, 0\}$ de M sont $\mathbf{V}$-ponctuelles, mais que $\{a, b\}$ ne l'est pas.

Dans la suite du problème, on prend pour $\mathcal{F}$ l'algèbre de Boole des langages de A^* reconnus par un monoïde de $\mathbf{V}$.

Question 8. Soit M un monoïde fini et soit $\pi : A^* \to M$ un morphisme surjectif. Soient p et q deux éléments de M. On suppose que les langages $\pi^{-1}(p)$ et $\pi^{-1}(q)$ sont séparables par un langage S de $\mathcal{F}$. Soit $\gamma : A^* \to N$ le morphisme syntactique de S et soit $\tau : M \to N$ le morphisme relationnel $\gamma \circ \pi^{-1}$. Montrer que $\tau(p)$ et $\tau(q)$ sont disjoints et en déduire que $\{p, q\}$ n'est pas un ensemble $\mathbf{V}$-ponctuel de M.

Question 9. Montrer réciproquement que si $\{p, q\}$ n'est pas un ensemble $\mathbf{V}$-ponctuel de M, alors les ensembles $\pi^{-1}(p)$ et $\pi^{-1}(q)$ sont $\mathcal{F}$-séparables.

Question 10. Montrer que le problème de la séparation pour $\mathcal{F}$ est équivalent au problème suivant: étant donné un monoïde fini M quotient de A^* et deux éléments $p, q \in M$, décider si la partie $\{p, q\}$ est $\mathbf{V}$-ponctuelle.

3. Langages testables par morceaux

Question 11. Soit $(u_n)_{n\geqslant 0}$ une suite de mots de A^* telle que pour tout n, u_n soit un sous-mot de u_{n+1}. Soit L un langage testable par morceaux de A^*. Montrer qu'il existe $N > 0$ tel que pour tout $n \geqslant N$, $u_n \in L$ ou pour tout $n \geqslant N$, $u_n \notin L$.

Soient K et L des langages disjoints. On appelle *zigzag* entre K et L une suite de mots $(u_n)_{n\geqslant 0}$ telle que pour tout $n \geqslant 0$,

(1) si n est pair, $u_n \in K$,
(2) si n est impair, $u_n \in L$,
(3) u_n est un sous-mot de u_{n+1}.

Question 12. On prend pour $\mathcal{F}$ l'ensemble des langages testables par morceaux de A^*. Montrer que K et L sont $\mathcal{F}$-séparables si et seulement si il n'existe aucun zigzag infini entre K et L.

4. Approche profinie.

On dit qu'un monoïde M de $\mathbf{V}$ *sépare* deux mots u et v s'il existe un morphisme $\varphi : A^* \to M$ tel que $\varphi(u) \neq \varphi(v)$. On fait l'hypothèse que deux mots distinct peuvent toujours être séparés par un monoïde de $\mathbf{V}$.

On pose, pour $u, v \in A^*$

$$r_{\mathbf{V}}(u, v) = \min\{|M| \mid M \in \mathbf{V} \text{ et } M \text{ sépare } u \text{ et } v\}$$
$$d_{\mathbf{V}}(u, v) = 2^{-r_{\mathbf{V}}(u,v)}$$

On rappelle (voir poly, Chapitre VII, section 2) que $d_{\mathbf{V}}$ est une distance ultramétrique, et que la complétion $\widehat{F}_{\mathbf{V}}(A)$ de A^* pour $d_{\mathbf{V}}$ est un monoïde compact. On note $\overline{L}$ la clôture d'une partie L de A^* dans $\widehat{F}_{\mathbf{V}}(A)$.

Question 13. Montrer que si K et L sont $\mathcal{F}$-séparables, alors $\overline{K}$ et $\overline{L}$ sont disjoints.

Question 14. Montrer réciproquement que si $\overline{K}$ et $\overline{L}$ sont disjoints, alors K et L sont $\mathcal{F}$-séparables.

Solution

1. Generalities and examples

Question 1. If L and L^c are $\mathcal{F}$-separable, then there exists a language $S \in \mathcal{F}$ such that $L \subseteq S$ and $S \cap L^c = \emptyset$. Thus necessarily $S = L$ and thus $L \in \mathcal{F}$.

Question 2. Easy direction. Let S be a language of $\mathcal{F}$ separating $\bigcup_{1 \leqslant i \leqslant n} K_i$ and $\bigcup_{1 \leqslant i \leqslant m} L_j$. Then S separates every pair K_i and L_j.

Let $S_{i,j}$ be a language of $\mathcal{F}$ containing K_i and disjoint from L_j. We claim that the language $S = \bigcup_{1 \leqslant i \leqslant n} \bigcap_{1 \leqslant j \leqslant m} S_{i,j}$ contains $\bigcup_{1 \leqslant i \leqslant n} K_i$ and is disjoint from $\bigcup_{1 \leqslant j \leqslant m} L_j$. Indeed, we have on the one hand $K_i \subseteq S_{i,j}$ for each $j \in J$ and thus $K_i \subseteq \bigcap_{1 \leqslant j \leqslant m} S_{i,j}$. It follows that $\bigcup_{1 \leqslant i \leqslant n} K_i \subseteq S$. On the other hand, we have $S \cap (\bigcup_{1 \leqslant j \leqslant m} L_j) = \bigcup_{1 \leqslant j \leqslant m} (S \cap L_j)$. Further, one gets

$$S \cap L_k = \bigcup_{1 \leqslant i \leqslant n} \Big((\bigcap_{1 \leqslant j \leqslant m} S_{i,j}) \cap L_k \Big) \subseteq \bigcup_{1 \leqslant i \leqslant n} (S_{i,k} \cap L_k) = \emptyset$$

which proves the claim. Since $\mathcal{F}$ is a Boolean algebra of languages, S belongs to $\mathcal{F}$.

Question 3. Let L be a star-free language containing $(a^2)^*$. Since L is star-free, there exists by Schützenberger's Theorem an integer n such that $a^n \sim_L a^{n+1}$. Since $a^{2n} \in (a^2)^*$, we get $a^{2n} \in L$ and thus $a^{2n+1} \in L$. Therefore L meets $(a^2)^*a$, a contradiction. The languages $(ab)^+$ and ba^* can be separated for instance by aA^*.

Question 4. Let S be a piecewise testable language. Then there exists $n > 0$ such that S is saturated by $\sim_n$. Now $(ab)^n \sim_n (ba)^n$ and thus if S contains $(ab)^+$, it will contain $(ab)^n$ and hence $(ba)^n$. It follows that S cannot separate $(ab)^+$ and $(ba)^+$.

The language a^* separates $(a^2)^*$ and $(b^2)^*b$. The syntactic monoid of a^* is idempotent and commutative and hence $\mathcal{J}$-trivial. Thus a^* is piecewise testable. Let F be any finite set of words of odd length. Then $a^* - F$ is also piecewise testable and also separates $(a^2)^*$ and $(b^2)^*b$. Suppose there is a minimal $\mathcal{F}$-separator for $(a^2)^*$ and $(b^2)^*b$. Then it should be contained in all languages $a^* - F$ and hence also in their intersection, which is equal to $(a^2)^*$. But $(a^2)^*$ is not piecewise testable.

Question 5. To separate $(ab)^+$ et $(ac)^+$, it suffices to take $S = A^*bA^*$. Let $S \in \mathcal{F}$ be such that $(a(b+c))^+ \subseteq S$ and $(ba)^+ \cap S = \emptyset$. Then $ab \in S$ by the first condition and thus $ba \in S$ since S is a commutative language. This yields a contradiction since $ba \in (ba)^+$.

Question 6. Since $\mathcal{F}$ is finite, it suffices to check for each language $S \in \mathcal{F}$ whether S separates L_1 and L_2.

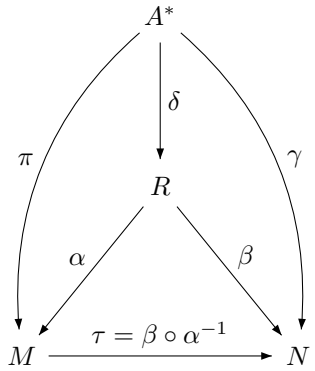
2. Pointlike sets

Question 7. Let $N \in \mathbf{V}$ and let $\tau : M \to N$ be a relational morphism. Let $n \in \tau(a)$. Since N is idempotent, $\{n\}$ is a subsemigroup of N and thus $\tau^{-1}(n)$ is a subsemigroup of M. Since $\tau^{-1}(n)$ contains a, it also contains $a^2 = 0$. It follows that $\{a, 0\} \subseteq \tau^{-1}(n)$ and thus $\{a, 0\}$ is $\mathbf{V}$-pointlike. Similarly, $\{b, 0\}$ is $\mathbf{V}$-pointlike.

Finally, let $\tau : M \to U_1$ be the relational morphism defined by $\tau(1) = \tau(b) = 1$, $\tau(a) = 0$ and $\tau(0) = \{0, 1\}$. Then $\tau^{-1}(0) = \{0, a\}$ and $\tau^{-1}(1) = \{1, b, 0\}$. Thus $\{a, b\}$ is not $\mathbf{V}$-pointlike.

Question 8. Let $P = \gamma(S)$. Since γ is the syntactic morphism of S, we also have $\gamma(S^c) = P^c$. Then the relations $\tau(p) = \gamma(\pi^{-1}(p)) \subseteq \gamma(S) = P$ and $\tau(q) = \gamma(\pi^{-1}(q)) \subseteq \gamma(S^c) = P^c$ show that $\tau(p)$ and $\tau(q)$ are disjoint. It follows that $\{p, q\}$ is not a $\mathbf{V}$-pointlike subset of M since if $\{p, q\} \subseteq \tau^{-1}(n)$, then $n \in \tau(p) \cap \tau(q)$.

Question 9. Suppose that $\{p, q\}$ is not a $\mathbf{V}$-pointlike subset of M. Then there is a relational morphism τ from M into a monoid N of $\mathbf{V}$ and an element n of N such that either $p \in \tau^{-1}(n)$ and $q \notin \tau^{-1}(n)$ or $q \in \tau^{-1}(n)$ and $p \notin \tau^{-1}(n)$. Let $\tau = \beta \circ \alpha^{-1}$ be the canonical factorization of τ. Since α is surjective, there is a morphism $\delta : A^* \to R$ such that $\pi = \alpha \circ \delta$. Setting $\gamma = \beta \circ \delta$, one has $\tau = \gamma \circ \pi^{-1}$ since $\gamma \circ \pi^{-1} = \beta \circ \delta \circ \pi^{-1} = \beta \circ \delta \circ (\alpha \circ \delta)^{-1} = \beta \circ \delta \circ \delta^{-1} \circ \alpha^{-1} = \beta \circ \alpha^{-1}$.



Let $S = \gamma^{-1}(n)$. Since $N \in \mathbf{V}$, the language L belongs to $\mathcal{F}$. Suppose that $p \in \tau^{-1}(n)$ and $q \notin \tau^{-1}(n)$. Since $\tau^{-1} = \pi \circ \gamma^{-1}$, one has $p \in \pi(L)$ and $q \notin \pi(L)$. It follows that L contains $\pi^{-1}(p)$ and is disjoint from $\pi^{-1}(q)$. The case where $q \in \tau^{-1}(n)$ and $p \notin \tau^{-1}(n)$ is symmetrical, since then L^c contains $\pi^{-1}(p)$ and is disjoint from $\pi^{-1}(q)$. It follows that the languages $\pi^{-1}(p)$ and $\pi^{-1}(q)$ are $\mathcal{F}$-separable.

Question 10. Let $\pi : A^* \to M$ be a surjective morphism. Suppose that the $\mathcal{F}$-separation problem is decidable. Then by Question 9, one can decide whether or not $\{p, q\}$ is pointlike.

Suppose that one can decide whether a two-element set is pointlike and let K and L be two regular languages of A^*. Then there exist a finite monoid M, a surjective morphism $\pi : A^* \to M$ and two subsets P and Q of M such that $K = \pi^{-1}(P)$ and $L = \pi^{-1}(Q)$. By Question 2, K and L are $\mathcal{F}$-separable if and only if, for each $p \in P$ and $q \in Q$, the languages $\pi^{-1}(p)$ and $\pi^{-1}(q)$ are $\mathcal{F}$-separable. By Questions 8 and 9, this is equivalent to saying that $\{p, q\}$ is not pointlike. It follows that $\mathcal{F}$-separation is decidable.

3. Piecewise testable languages

Question 11. Let S be a piecewise testable language. Then there exists $k > 0$ such that S is saturated by $\sim_k$. Denote by $S_k(u)$ the set of subwords of length $\leqslant k$ of u. If $(u_n)_{n \geqslant 0}$ is a sequence of words such that, for each n, u_n is a subword of u_{n+1}, then $S_k(u_n) \subseteq S_k(u_{n+1})$ and since there are only finitely many words of length $< k$, there exists N such that for all $n \geqslant N$, $S_k(u_n) = S_k(u_N)$, it follows that $u_n \sim_k u_N$ and thus either for all $n > N$, $u_n \in L$ or for all $n > N$, $u_n \notin L$.

Question 12. Suppose there exists an infinite zigzag $(u_n)_{n \geqslant 0}$ between K and L and let S be a piecewise testable langage separating K and L. Then by construction, $u_n \in S$ if n is even and $u_n \notin S$ if n is odd, a contradiction with question 11.

Conversely,

4. A profinite approach

Question 13. Suppose that K and L are separated by a language S of $\mathcal{F}$. Then $\overline{K} \cap \overline{L} \subseteq \overline{S} \cap \overline{S^c}$. But in the pro-$\mathbf{V}$ topology, $\overline{S^c} = (\overline{S})^c$ and thus $\overline{S} \cap \overline{S^c} = \emptyset$. Therefore $\overline{K} \cap \overline{L} = \emptyset$.

Question 14. Suppose that $\overline{K} \cap \overline{L} = \emptyset$. Then every $u \in \overline{K}$ belongs to the open set $(\overline{L})^c$. Since the closure of the languages of $\mathcal{F}$ form a basis of the topology of $\widehat{F}_{\mathbf{V}}(A)$, there exists a language S_u of $\mathcal{F}$ containing u such that $\overline{S_u} \subseteq (\overline{L})^c$. Thus $\overline{K} \subseteq \bigcup_{u \in \overline{K}} \overline{S_u}$. Since $\overline{K}$ is closed, hence compact, one can extract a finite cover $\overline{S_{u_1}}, \ldots, \overline{S_{u_n}}$. Setting $S = S_{u_1} \cup \cdots \cup S_{u_n}$, we get $\overline{K} \subseteq \overline{S}$, whence $K \subseteq S$ since $\overline{K} \cap A^* = K$ and $\overline{S} \cap A^* = S$. Moreover $S \subseteq \overline{S} \subseteq (\overline{L})^c \subseteq L^c$.