Master Parisien de Recherche en Informatique

Course 2.16 – Finite automata based computation models

7 march 2022 — Exam — Probabilistic Automata

Books and computers forbidden — Lecture and personal notes allowed.
This part should be written on separate test papers.

Notations and terminology. For any set X of words, X^* denotes the Kleene star of X. For any alphabet A and any word $w \in A^*$, we denote by $|w|$ the length of w, and $|w|_a$ the number of times the letter $a \in A$ appears in w. For example, if $A = \{a, b, c\}$ and $w = abbcab$ then $|w| = 6$, $|w|_a = 2$, $|w|_b = 3$ and $|w|_c = 1$. We use the usual convention that $x^0 = 1$ for all nonzero $x \in \mathbb{R}$. Recall that for a probabilistic automaton $\mathcal{A}$ and an isolated cut-point λ, we say that the *isolation threshold* of λ is the largest $\delta > 0$ such that $|\mathcal{A}(w) - \lambda| \geqslant \delta$ for all words w.

Deterministic versus probabilistic automata

Let $m \geqslant 1$ and $A = \{a_1, \ldots, a_m\}$ be an alphabet of size m. We consider the language L_m over A that contains each letter of the alphabet exactly m times:

$$L_m = \{w \in A^* : \forall a \in A, |w|_a = m\}.$$

(a1) By using Myhill-Nerode theorem, show that the smallest deterministic complete finite automaton recognizing L_m has exactly $(m+1)^m + 1$ states.

Let $\Sigma = \{x\}$ be a unary alphabet. For any $n \in \mathbb{N}$, let $C_n = \{x^n\}$ be the language over Σ consisting of a single word x^n. We admit the following result, proven in the (exercise sheet of the) course.

Lemma 1. *There exists $\delta > 0$ such that for every n, there exists a probabilistic automaton $\mathcal{A}_n$ with $O\left(\frac{\ln^2 n}{\ln \ln n}\right)$ states that recognizes C_n with an isolated cut-point and isolation threshold at least δ.*

(a2) Explain why we can assume that the cut-point Lemma 1 is $9/10$, *i.e.* $C_n = \mathcal{L}_{\mathcal{A}_n}(9/10)$. You will need to justify that this cut-point is still isolated with constant threshold δ' independent of n.

Let p be a prime greater than αm for some $\alpha \in \mathbb{N}$ to fix later. For any $i \in \{1, \ldots, \alpha m\}$ and $v \in \mathbb{N}$, consider the language

$$L'_{i,v} = \left\{ w \in A^* : \sum_{t=1}^{m} (i^{t-1} \bmod p)|w|_{a_t} = v \right\}.$$

(a3) By using Lemma 1, show that for any i and v, there exists a probabilistic automaton $\mathcal{B}_{i,v}$ with $O\left(\frac{\ln^2 v}{\ln \ln v}\right)$ states that recognizes $L'_{i,v}$ with isolated cut-point $9/10$ and isolation threshold at least δ. *Hint: reading one letter in $\mathcal{B}_{i,v}$ corresponds to reading several letters at once in $\mathcal{A}_v$.*

(a4) Show that for every $i \in \{1, \ldots, \alpha m\}$, $L_m \subseteq L'_{m,i,v_i}$ for a certain value $v_i \in \{0, \ldots, m^2 p\}$ that you will identify.

We admit the following lemma, whose proof is deferred to question **(a10)**

Lemma 2. *Let $y_1, \ldots, y_m \in \{1, \ldots, \alpha m\}$ be pairwise distinct, then the vectors $z_0, \ldots, z_{m-1}$ defined by*

$$z_j = (y_1^j \bmod p, y_2^j \bmod p, \ldots, y_m^j \bmod p)$$

are linearly independent.

(a5) Let $S \subseteq \{1, \ldots, \alpha m\}$. Show that if $|S| \geqslant m$ then $\bigcap_{i \in S} L'_{m,i,v_i} \subseteq L_m$ using Lemma 2.

We now consider the probabilistic automaton $\mathcal{C}_m = \frac{1}{\alpha m} \sum_{i=1}^{\alpha m} \mathcal{B}_{i,v_i}$ where v_i is defined as in question **(a4)**.

(a6) Show that if $w \in L_m$ then $\mathcal{C}(w) \geqslant 9/10 + \delta$.

(a7) Show that if $w \notin L_m$ then $\mathcal{C}(w) \leqslant \frac{9}{10} - \delta + \frac{1/10+\delta}{\alpha}$.

(a8) Show that there exists a choice of α, independent of m, such that $\mathcal{C}_m$ recognizes L_m with an isolated cut-point and isolation threshold at least $\delta/2$. Show that $\mathcal{C}_m$ has $O\left(m \frac{\ln^2 n}{\ln \ln n}\right)$ states. *Hint: you can use the fact that we can choose p such that $p = \alpha n + o(\alpha n)$.*

(a9) Show that there exists $\delta > 0$ such that for infinitely many n, there exists a regular language recognized by a probabilistic automaton with n states and an isolated cut-point with isolation threshold at least δ, such that the smallest deterministic finite automaton recognizing it has $\Omega(2^{\frac{n \ln \ln n}{\ln n}})$ states. Compare with the result in the course about isolated cut-points.

(a10) Let $y_1, \ldots, y_m$ be as in Lemma 2 and assume that $z_1, \ldots, z_m$ are linearly dependent. Show that there exists $c_0, \ldots, c_{m-1}$ not all zero such that $c_0 + c_1 x + \cdots + c_{m-1} x^{m-1} = 0 \bmod p$ for all $x \in \{y_1, \ldots, y_m\}$. Prove Lemma 2. *Hint: you can use the fact that a degree d polynomial with integer coefficients has at most d distinct roots modulo any prime number $p > d$.*

References

[Amb96] Andris Ambainis. The complexity of probabilistic versus deterministic finite automata. In *Proceedings of the 7th International Symposium on Algorithms and Computation*, ISAAC '96, page 233–238, Berlin, Heidelberg, 1996. Springer-Verlag.

Solutions to exercises

(a1) Intuitively, the automaton needs to count each of the m letters up to m, and as soon as one goes above m, we can reject. Hence we need $(m+1)^m$ states to count $\{0,\dots,m\}^m$, and one extra state to reject.

Let $\equiv_{L_m}$ denote the Myhill-Nerode equivalence relation for L_m. For any $k_1,\dots,k_m \in \mathbb{N}$, define $w(k_1,\dots,k_m) = a_1^{k_1}\cdots a_m^{k_m}$. Let $(k_1,\dots,k_m) \neq (k'_1,\dots,k'_m) \in \{0,\dots,m\}^m$, then $w(k_1,\dots,k_m) \not\equiv_{L_m} w(k'_1,\dots,k'_m)$. Indeed, on the one hand we have $w(k_1,\dots,k_m)w(m-k_1,\dots,m-k_m) \in L_m$ since each letter a_i appears $k_i + m - k_i = m$ times (note that we used that $k_i \leqslant m$ for $m - k_i$ to be nonnegative). On the other hand, there is i such that $k_i \neq k'_i$ and therefore the word $w(k'_1,\dots,k'_m)w(m-k_1,\dots,m-k_m) \notin L_m$ because it contains $k'_i + m - k_i \neq m$ times the letter a_i. Furthermore, for any $(k_1,\dots,k_m) \in \{0,\dots,m\}^m$, $w(k_1,\dots,k_m) \not\equiv_{L_m} a_1^{m+1}$. Indeed, we have seen that $w(k_1,\dots,k_m)w(m-k_1,\dots,m-k_m) \in L_m$ but $a_1^{m+1}w(m-k_1,\dots,m-k_m) \notin L_m$ because the letter a_1 appears at least $m+1+m-k_1 > m$ times (since $k_1 \leqslant m$).

We have therefore shown that $\equiv_{L_m}$ has *at least* $(m+1)^m + 1$ equivalences classes ($(m+1)^m$ is the cardinal of $\{0,\dots,m\}^m$ and the $+1$ is for the word a_1^{m+1}). By the Myhill-Nerode theorem, any DFA that recognizes L_m has at least that many states On the other hand, it is trivial to build a DFA with $(m+1)^m+1$ states that recognizes L_m by counting the number of each letters up to m and adding one extra state to reject as soon as a letter appears $> m$ times.

(a2) Let $\mathcal{A}_n$ be the automaton of the lemma, then $C_n = \mathcal{L}_{\mathcal{A}_n}(\lambda_n)$ for some λ_n. Clearly $\lambda_n \neq 0, 1$ because C_n is not the empty language, nor the universal one. There are two cases:

- If $\lambda_n \geqslant 9/10$ then we can let $\mathcal{A}'_n = \frac{9}{10\lambda_n}\mathcal{A}_n$ by multiplying the probability of the intial states of $\mathcal{A}_n$ by $\frac{9}{10\lambda_n} \in [0,1]$. We immediately have that $\mathcal{L}_{\mathcal{A}'_n}(9/10) = \mathcal{L}_{\mathcal{A}_n}(\lambda_n) = C_n$. Furthermore, for all $w \in A^*$,

$$|\mathcal{A}'_n(w) - \tfrac{9}{10}| = |\tfrac{9}{10\lambda_n}\mathcal{A}_n(w) - \tfrac{9}{10}| = \tfrac{9}{10\lambda_n}|\mathcal{A}_n(w) - \lambda_n| \geqslant \tfrac{9\delta}{10\lambda_n} \geqslant \tfrac{9}{10}\delta$$

  since $\lambda_n \leqslant 1$, and is therefore independent of n.

- If $\lambda_n < 9/10$ then we can let $\mathcal{A}'_n = \alpha\mathcal{A}_n + (1-\alpha)$, where $\alpha = \frac{1-9/10}{1-\lambda_n} = \frac{1}{10(1-\lambda_n)}$, by doing a convex combination with the automata that accepts all words. Note there that $\alpha \in [0,1]$ because $0 < \lambda_n < 9/10$. A small computation shows that $\mathcal{L}_{\mathcal{A}'_n}(9/10) = \mathcal{L}_{\mathcal{A}_n}(\lambda_n) = C_n$. Furthermore, for all $w \in A^*$,

$$|\mathcal{A}'_n(w) - \tfrac{9}{10}| = |\alpha\mathcal{A}_n(w) + (1-\alpha) - \tfrac{9}{10}| = \alpha|\mathcal{A}_n(w) - \lambda_n| \geqslant \alpha\delta \geqslant \tfrac{\delta}{10}$$

  since $\lambda_n \geqslant 0$, and is therefore independent of n.

In summary, we have shown that the isolation threshold is always at least $\frac{\delta}{10}$ which is independent of n.

(a3) We consider the automaton $\mathcal{B}_{i,v}$ that has the same states as $\mathcal{A}_v$ (including the same initial and final states). We modify the transitions so that for any pair of states q, q' and letter $a_t \in A$,

$$\mathbb{P}_{\mathcal{B}_{i,v}}\left(q \xrightarrow{a_t} q'\right) = \mathbb{P}_{\mathcal{A}_v}\left(q \xrightarrow{x^{\ell_{a_t}}} q'\right) \text{ where } \ell_{a_t} = i^{t-1} \bmod p.$$

Technically, this can be done by defining the transition matrix of a_t in $\mathcal{B}_{i,v}$ to be equal to $\mu^{\ell_{a_t}}$ where μ is the transition matrix of $\mathcal{A}_v$. In other words, reading a_t in $\mathcal{B}_{i,v}$ is like reading $x^{\ell_{a_t}}$ in $\mathcal{A}_v$. Note that ℓ_{a_t} only depends on t (and i is fixed) and is positive (since $p \nmid i^{t-1}$ by primality of p and the fact that $i \leqslant \alpha m < p$) so this is well-defined. Now given a word $w \in A^*$, it follows that the probability of acceptance of w is $\mathcal{B}_{i,v}(w) = \mathcal{A}_v(x^M)$ where

$$M = \sum_{k=1}^{|w|} \ell_{w_k} = \sum_{t=1}^{m} \ell_{a_t}|w|_{a_t} = \sum_{t=1}^{m} (i^{t-1} \bmod p)|w|_{a_t}.$$

Again, technically, this can be shown by using the matrix definition above (call S and T the initial and final vectors of both $\mathcal{A}_v$ and $\mathcal{B}_{i,v}$):

$$\mathcal{B}_{i,v}(w) = S\mu^{\ell_{w_1}}\cdots\mu^{\ell_{w_{|w|}}}T = S\mu^{\sum_{i=1}^{|w|}\ell_{w_i}}T = \mathcal{A}_v(x^M).$$

Finally, we conclude by the fact that $\mathcal{A}_v$ only recognizes those words x^M such that $M = v$. Note that this construction has the same cut-point and isolation threshold as $\mathcal{A}_v$. By question **(a2)**, we can assume that the $\mathcal{A}_v$ have cut-point 9/10.

(a4) If $w \in L_m$ then $|w|_{a_t} = m$ for all t. Therefore, for all i,

$$\sum_{t=1}^{m} (i^{t-1} \bmod p)|w|_{a_t} = m \sum_{t=1}^{m} (i^{t-1} \bmod p).$$

Hence if we let v_i be the right-hand side, we indeed have that $w \in L'_{i,v_i}$. We finally check that

$$v_i = m \sum_{t=1}^{m} (i^{t-1} \bmod p) \leqslant m^2 p.$$

(a5) We will show the result for $|S| = m$. This will imply the result for all $|S| \geqslant m$ since having more elements only makes the intersection smaller. Denote the elements of S by $y_1, \ldots, y_m$. If $w \in \bigcap_{i \in S} L'_{m,i,v_i}$ then, by definition, $\sum_{t=1}^{m} (i^{t-1} \bmod p)|w|_{a_t} = v_i = m \sum_{t=1}^{m} (i^{t-1} \bmod p)$ for all $i \in S$. Therefore

$$\sum_{t=1}^{m} (|w|_{a_t} - m)(i^{t-1} \bmod p) = 0$$

for all $i \in S$. Using the notation of Lemma 2, this can be written as $\sum_{t=1}^{m} (|w|_{a_t} - m)(z_t)_j = 0$ for all $j \in \{1, \ldots, m\}$ since $S = \{y_1, \ldots, y_m\}$. Therefore $\sum_{t=1}^{m} (|w|_{a_t} - m) z_t = 0$. But the y_i are pairwise distinct by definition, so by Lemma 2, $z_0, \ldots, z_{m-1}$ are linearly independent, hence $|w|_{a_t} - m = 0$ for all t. This shows that $w \in L_m$.

(a6) If $w \in L_m$ then $w \in L'_{i,v_i}$ for all $i = 1, \ldots, \alpha m$, by question **(a4)**. By question **(a3)**, $\mathcal{B}_{i,v_i}$ recognizes L'_{i,v_i} so $\mathcal{B}_{i,v_i}(w) \geqslant 9/10 + \delta$ since the cut-point has isolation threshold δ. By construction of $\mathcal{C}_m$, it immediately follows that $\mathcal{C}_m(w) \geqslant 9/10 + \delta$.

(a7) Let $w \notin L_m$ and let $S = \{i : w \in L'_{i,v_i}\} \subseteq \{1, \ldots, \alpha m\}$. By question **(a5)**, $|S| < m$ for otherwise we would have $w \in L_m$. For $i \in S$, we have $\mathcal{B}_{i,v_i} \leqslant 1$ since it is a probability. But since $\mathcal{B}_{i,v_i}$ has isolation threshold δ by question **(a3)**, if $i \notin S$, then $\mathcal{B}_{i,v_i} \leqslant 9/10 - \delta$. Therefore,

$$\begin{aligned}
\mathcal{C}(w) &= \frac{1}{\alpha m} \sum_{i=1}^{\alpha m} \mathcal{B}_{i,v_i}(w) \\
&= \frac{1}{\alpha m} \left(\sum_{i \in S} \mathcal{B}_{i,v_i}(w) + \sum_{i \notin S} \mathcal{B}_{i,v_i}(w) \right) \\
&\leqslant \frac{1}{\alpha m} \left(|S| + |\{1, \ldots, \alpha m\} \setminus S| (\tfrac{9}{10} - \delta) \right) \\
&\leqslant \frac{1}{\alpha m} \left(m + (\alpha - 1) m (\tfrac{9}{10} - \delta) \right) \\
&= \frac{9}{10} - \delta + \frac{1/10 + \delta}{\alpha}.
\end{aligned}$$

(a8) It suffices to choose α such that $-\delta + \frac{1/10+\delta}{\alpha} \leqslant 0$ which is always possible because $\frac{1/10+\delta}{\alpha} \to 0$ as $\alpha \to \infty$. Note that this choice does not depend on m. The number of states of $\mathcal{C}$ is the sum of the number of states of the $\mathcal{B}_{i,v_i}$ for $i = 1, \ldots, \alpha m$. Automaton $\mathcal{B}_{i,v_i}$ has as many states as $\mathcal{A}_{v_i}$ which is $O\left(\frac{\ln^2 v_i}{\ln \ln v_i}\right)$. On the other hand, $v_i \leqslant m^2 p$ by question **(a4)**. By the distribution of primes, we can always choose $p = \alpha m + o(\alpha m)$ and α was chosen to be a constant that only depends on δ and is independent of m. Therefore $v_i = O(m^3)$ and $\mathcal{C}$ has

$$\alpha m \cdot O\left(\frac{\ln^2 O(m^3)}{\ln \ln O(m^3)} \right) = O\left(m \frac{\ln^2 m}{\ln \ln m} \right)$$

states.

(a9) Putting questions **(a1)** and **(a8)** together, for every m, we have found a language L_m recognized by a probabilistic automaton with $n = O\left(m \frac{\ln^2 m}{\ln \ln m}\right)$ states, but whose smallest DFA that recognizes it has $N = (m+1)^m$ states. First observe that $N = (m+1)^m = 2^{O(m \ln m)}$ and that $n \frac{\ln \ln m}{\ln m} = O(m \ln m)$. Furthermore, observe that

$$n = O\left(m \frac{\ln^2 m}{\ln \ln m} \right) \;\Rightarrow\; n = \Omega(m) \text{ and } n = O(m^2) \;\Rightarrow\; \ln n = \Theta(\ln m) \;\Rightarrow\; \ln \ln n = \Theta(\ln \ln m).$$

It follows that

$$N = 2^{O(n\frac{\ln\ln m}{\ln m})} = 2^{O(n\frac{\ln\ln n}{\ln n})}.$$

By comparison, the result from the lecture says that for a cut-point language with isolation threshold δ, the number of states for DFA is bounded by $(1+\frac{r}{\delta})^{n-1}$ where r is the number of accepting states. Clearly r is smaller than the number of states which is $O(m^2)$, and recall that δ is constant, therefore the upper bound of the theorem is

$$2^{(n-1)\ln(1+\frac{r}{\delta})} = 2^{O(n\ln m)} = 2^{O(n\ln n)}.$$

Therefore there is a still a gap between this upper bound and what we obtain but the two bounds are quite close.

(a10) If there are linearly dependent, there exists $c_0,\dots,c_{m-1}$ such that $c_0 z_0 + \dots + c_{m-1}z_{m-1} = 0$. Therefore, for all t,

$$0 = (c_0 z_0 + \dots + c_{m-1}z_{m-1})_t = c_0 y_t^0 + c_1 y_t^1 \cdots + c_{m-1}y_t^{m-1} \bmod p.$$

Since $y_1,\dots,y_m \in \{1,\dots,\alpha m\}$ then in particular $y_i < p$ so the y_i are pairwise distinct *modulo* p. Therefore the polynomial $P(x) = c_0 + c_1 x + \dots + c_{m-1}x^{m-1}$, which has degree at most $m-1$, has at least m distinct roots modulo p. This is a contradiction with the hint since $p > m$.

Note: the hint can be proven by induction on the degree of P. If P has degree 1 then $P(x) = a + bx$ for some a and $b \neq 0 \bmod p$ (otherwise this is trivial). If x, y are such that $P(x) = P(y) = 0 \bmod p$ then $a + bx = a + by \bmod p$ so $x = y \bmod p$ (b is invertible modulo p, by primality of p) so P has only one root modulo p. Now if P has degree $d > 1$, assume that P has at least one root modulo p (otherwise the result is proved already): $P(x_0) = 0 \bmod p$ for some x_0. Then we can write $P(x) = (x - x_0)Q(x) \bmod p$ for some polynomial Q of degree $d-1$ (simply consider the expansion of $P(x_0 + x) \bmod p$ to find Q). But now, if y is such that $y \neq x_0 \bmod p$ and $P(y) = 0 \bmod p$ then it must be the case that $Q(y) = 0 \bmod p$ (again by primality of p). By induction, Q has at most $d-1$ solutions modulo p, therefore there can only be d roots of P modulo p.

This can also be shown more abstractly: any nonzero polynomial $P \in R[x]$ of degree d, where R is an (integral) domain, has at most d roots in R. In fact this is a characterization of integral domains.