Tree Languages and Applications

M1 Informatique – ENS Paris-Saclay

Exam, January 12, 2023

Time : two hours. All answers must come with a justification. Results from the course can of course be used without proof.

1 Residuals

For $\mathcal{F} = \{f(2), a(0)\}$ and $n > 0$, let L_n be the language of trees that have at least one branch of length exactly n, i.e.

$$L_n = \{\, t \in T(\mathcal{F}) \mid \exists p \in Pos(t) : |p| = n - 1 \wedge t(p) = a \,\}.$$

E.g., $f(a, f(a, f(a, a))) \in L_3$ because it contains one branch of length 3 (as well as one of length 2 and two of length 4).

(a) Give a (bottom-up) NFTA for L_n with $n + 1$ states.

(b) Show that the minimal DFTA for L_n has at least 2^{n-1} states.

Let $L \subseteq T(\mathcal{F})$ be a language of trees and $\mathcal{C} \in \mathcal{C}(\mathcal{F})$ a context. The *residual* of L by C is defined as $C^{-1}L := \{\, t \in T(\mathcal{F}) \mid C[t] \in L \,\}$. We define $R(L) = \{\, C^{-1}L \mid C \in \mathcal{C}(\mathcal{F}) \,\}$ as the set of residuals of L.

(c) Show that if L is recognizable, then $|R(L)|$ is finite.

(d) Show that for L_n as above, $|R(L_n)| = n + 2$.

2 Prime decompositions

Let $\mathcal{F} = \{0(1), 1(1), \bot(0)\}$. For $n \in \mathbb{N}$, its encoding $\widetilde{n}$ is defined as :

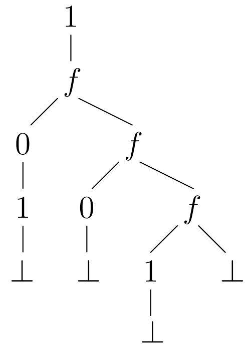
- $\widetilde{0} = 0(\bot)$ and $\widetilde{1} = 1(\bot)$;
- if $n = 2m > 0$, then $\widetilde{n} = 0(\widetilde{m})$;
- if $n = 2m + 1 > 1$, then $\widetilde{n} = 1(\widetilde{m})$.

In other words, $\widetilde{n}$ is the (reverse) binary encoding of n, without leading zeros.

Moreover, let $\mathcal{F}' = \{\, \langle f, g, h\rangle(k) \mid f \in \mathcal{F}_m, g \in \mathcal{F}_n, h \in \mathcal{F}_\ell, k = \max\{m, n, \ell\} \,\}$. A tree over $\mathcal{F}'$ encodes a triple of natural numbers, with $\bot$ filling unused positions, e.g., $\langle \widetilde{2}, \widetilde{1}, \widetilde{5}\rangle = \langle 011\rangle(\langle 1\bot 0\rangle(\langle \bot\bot 1\rangle(\langle \bot\bot\bot\rangle)))$.

(a) Show that $L = \{\, \langle \widetilde{n}, \widetilde{m}, \widetilde{n+m}\rangle \mid n, m \in \mathbb{N} \,\}$ is recognizable. Give an accepting run of your automaton on $\langle \widetilde{6}, \widetilde{3}, \widetilde{9}\rangle$.

We now consider another encoding $\overline{n}$ for $n \in \mathbb{N}$, using trees over $\mathcal{G} = \{0(1), 1(1), \perp(0), f(2)\}$. If $n > 1$, let $p_1, \ldots, p_k$ be the (unique) increasing sequence of prime numbers up to p_k, where p_k is the largest prime factor of n. There are $n_1, \ldots, n_k$ such that $n = \prod_{i=1}^{k} p_i^{n_i}$. Then we let $\overline{n} = 1(f(\widetilde{n_1}, f(\widetilde{n_2}, \ldots, f(\widetilde{n_k}, \perp) \ldots)))$. Moreover, define $\overline{0} = 0(\perp)$ and $\overline{1} = 1(\perp)$. E.g., $\overline{20}$ is shown below, given that $20 = 2^2 \cdot 3^0 \cdot 5^1$:



(b) Show that $\{ \overline{n} \mid n \in \mathbb{N} \}$ is recognizable.

(c) Show that $\{ \langle \overline{n}, \overline{m}, \overline{n \times m} \rangle \mid n, m \in \mathbb{N} \}$ is recognizable.

3 Closures

Let $\mathcal{F} = \{f(2)\} \cup \Sigma$, where $\Sigma = \{a, b\}$. For $t \in T(\mathcal{F})$, let $fr(t) \in \Sigma^*$ denote the word obtained from reading the leaves of t from left to right, i.e. in increasing lexicographical order of their positions.

We call $L \subseteq T(\mathcal{F})$ *closed under commutativity* if $C[f(t, t')] \in L$ implies $C[f(t', t)] \in L$, for any context $C \in \mathcal{C}(\mathcal{F})$ and trees $t, t' \in T(\mathcal{F})$. We call L *closed under associativity* if $C[f(t, f(t', t''))] \in L$ implies $C[f(f(t, t'), t'')] \in L$ and vice versa. The *closure* of some $L \subseteq T(\mathcal{F})$ under commutativity/associativity is the least tree language containing L and closed under commutativity/associativity.

(a) Let $L_1 \subseteq T(\mathcal{F})$ be the language of trees having the same number of a-leaves as b-leaves. Is L_1 recognizable ?

(b) Let $L_2 \subseteq T(\mathcal{F})$ be the least set of trees containing $f(a, b)$ and such that $t \in L_2$ implies $f(f(a, t), b)) \in L_2$. Is L_2 recognizable ?

(c) Let $L \subseteq \Sigma^*$ be a regular *word* language. Is the tree language $\{ t \in T(\mathcal{F}) \mid fr(t) \in L \}$ recognizable in general ?

(d) Let $L \subseteq T(\mathcal{F})$ recognizable. Is the associative closure of L recognizable in general ?

(e) Let $L \subseteq T(\mathcal{F})$ recognizable. Is the associative and commutative closure of L recognizable in general ?

(f) Let $L \subseteq T(\mathcal{F})$ recognizable. Is the commutative closure of L recognizable in general ?