Tree Languages and Applications

M1 Informatique – ENS Paris-Saclay

Sample solutions for the exam (January 9, 2021)

1 Recognizable tree languages

Question (a)

L_1 is recognizable by an NFTA $\langle\{q_a, q_b, q_f, q_r\}, \mathcal{F}, \{q_f\}, \Delta\rangle$ with the following rules:

$$a \to q_a \qquad b \to q_b \qquad f(q_a, q_b) \to q_f \qquad f(q_f, q_b) \to q_r \qquad f(q_a, q_r) \to q_f$$

Question (b)

L_2 is not recognizable. To see this, consider the language L_3 of trees where a occurs to the left of the root and b to its right. L_3 is recognized by the NFTA $\langle\{q_a, q_b, q_f\}, \mathcal{F}, \{q_f\}, \Delta\rangle$ with

$$a \to q_a \qquad f(q_a, q_a) \to q_a \qquad b \to q_b \qquad f(q_b, q_b) \to q_b \qquad f(q_a, q_b) \to q_f$$

If L_2 was recognizable, then so would $L_4 := L_2 \cap L_3$ be. But L_4 contains

$$f(\underbrace{f(f(\cdots(f(}_{i}a, a), \ldots), a), a), \underbrace{f(f(\cdots(f(}_{i}b, b), \ldots), b), b))$$

for every $i \geq 0$. It is now trivial to apply the pumping lemma to show that L_4 is not recognizable. But then neither is L_2.

Question (c)

If L is a recognizable word language, then it is recognized by a morphism ϕ from monoid $\langle M, \cdot\rangle$. Then the corresponding tree language is recognized by DFTA $\langle M, \mathcal{F}, \phi(L), \Delta\rangle$, where Δ contains $a \to \phi(a)$, $b \to \phi(b)$, and $f(m, m') \to m \cdot m'$ for all $m, m' \in M$.

Question (d)

Let $\mathcal{A} = \langle Q, \mathcal{F}, G, \Delta\rangle$ an NFTA recognizing L. Then $\mathcal{A}' = \langle Q, \mathcal{F}, G, \Delta \cup \Delta'\rangle$ recognizes the commutative closure of L, where $\Delta' = \{\, f(q, q') \to q'' \mid f(q', q) \to q'' \in \Delta \,\}$.

Question (e)

L_1 is recognizable. The commutative and associative closure of L_1 is L_2, which is not recognizable.

Question (f)

Let L_3 and L_4 as in the proof of (b). Again, L_1 is recognizable, but the associative closure of L_1 intersected with L_3 is L_4, which is not recognizable.

2 Tree relations

Question (a)

R can equivalently be written as $\{ \langle f^i(a), f^{i+1}(a) \rangle \mid i \geq 0 \}$.

(i) Yes, see (iii), with $\mathfrak{T}_2 \subseteq \mathfrak{R}_2$.

(ii) No. Suppose that R was a finite union of cross-products of the form $\bigcup_{i=1}^{n} (L_1^i \times L_2^i)$. But for every tree $t = f^i(a)$, $i \geq 1$, there is only one t' with $\langle t, t' \rangle \in R$ and only one t'' with $\langle t'', t \rangle \in R$. So all the $L_1^{(i)}, L_2^{(i)}$ must be singletons, and the finite union above would only contain finitely many pairs, but R is infinite.

(iii) Yes; it is easy to construct NFTA $\mathcal{A}_1, \mathcal{A}_2$ with one single common final state and $\mathcal{L}(\mathcal{A}_1) = \{a\}$ and $\mathcal{L}(\mathcal{A}_2) = \{f(a)\}$.

Question (b)

No. Take $\mathcal{F} = \{g(2), a, b, c\}$. Let R_1 be the relation that allows to replace a-leaves by b, and R_2 the relation that allows to replace a by c. A GTT accepting the union would have to accept the pairs $\langle a, b \rangle, \langle a, c \rangle$ and therefore also the pair $g(a, a), g(b, c)$, which is not in $R_1 \cup R_2$.

Question (c)

No, because every GTT-recognizable relation contains at least the identity relation.

3 Unranked trees

Question (a)

Let us call a state in an NHA or NFA *productive* if it is part of an accepting run. I.e., in an NFA a state is productive if it is reachable from an initial state and can reach a final state. And in a NHA a state q is productive if there exists a context C and a ground tree t such that $t \to^* q$ and $C[q]$ is accepted by the NHA.

Let $\mathcal{A} = \langle Q, \Sigma, G, \Delta \rangle$ be a NHA(NFA). As in the finiteness problem for NFTA, the language accepted by $\mathcal{A}$ is infinite iff there is some productive state $q \in Q$ with a "loop", e.g. such that $C'[q] \to^* q$ for some non-trivial context C'. (Indeed, if no such state existed, then $\mathcal{A}$ could only accept trees of finite height.)

It is known that non-productive states can be removed from an NFA in polynomial time, by simple reachability analysis. So w.l.o.g. one can assume that the horizontal NFA of $\mathcal{A}$ contain only productive states. We then construct the directed graph $\mathcal{G}$ whose nodes are Q and that contains an edge $\langle q, q' \rangle$ iff Δ contains some rule $a(R) \to q'$ where the NFA for R contains a transition reading q. Thus, there is a path from q to q' in $\mathcal{G}$ iff there exists a context C' s.t. $C'[q] \to^* q'$.

We will call a state $q \in Q$ *initial* if there exists some rule $a(R) \to q$ with $\varepsilon \in R$. Now $\mathcal{A}$ accepts an infinite language iff $\mathcal{G}$ contains a state q reachable from an initial state, reaching a final state, and with a non-empty loop around itself. This can clearly be tested in polynomial time (indeed, linear time if one uses SCC decomposition.)

Question (b)

$(a(ba)^*c)^+$

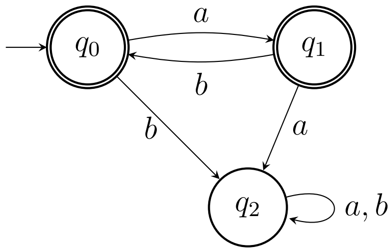
Question (c)

The expression means that the parts to the left and right of # must have some letter in common. This is also achieved by the following:

$$\big(aa^*(\#b^*a\Sigma^* + b\Sigma^*\#\Sigma^+)\big) + \big(bb^*(\#a^*b\Sigma^* + a\Sigma^*\#\Sigma^+)\big)$$

Question (d)

The simplest expression I could come up with is $(ab)^*(a + \varepsilon)$, with the following minimal complete DFA:



Languages recognizable by deterministic RE are characterized in a paper by Brüggemann-Klein and Wood (BKW98 in the bibliographic notes of the TATA book, which can be found online) – congratulations to Luc for finding it!

Theorem E in BKW98, with the definitions immediately before it, tells us that this property is related to the outgoing transitions of the strongly connected components ("orbits") of the minimal DFA recognizing the language.