Tree Automata and Applications
Exercise session 1 Solutions

Luc Lapointe
luc.lapointe@ens-paris-saclay.fr
home.lmf.cnrs.fr/LucLapointe/

Exercise 1 - First constructions of Tree Automata

Let $\mathcal{F} = \{f(2), g(1), a(0)\}$. Give a DFTA and a top-down DFTA for the set $G(t)$ of ground instances of the term $t = f(f(a,x), g(y))$ which is defined by:

$$G(t) = \left\{f(f(a,u), g(v)) \mid u, v \in T(\mathcal{F})\right\}$$

Detailed solution

Top-down DFTA: $Q = \left\{q_{f_1}, q_{f_2}, q_a, q_g, q_\top\right\}$, $I = \left\{q_{f_1}\right\}$, and transitions as follow:

$$q_{f_1}(f(x,y)) \to f\left(q_{f_2}(x), q_g(y)\right) \text{ (i)} \qquad q_{f_2}(f(x,y)) \longrightarrow f\left(q_{a(x)}, q_{\top(y)}\right) \text{ (ii)}$$

$$q_g(g(x)) \longrightarrow g(q_\top(x)) \quad \text{(iii)} \qquad q_a(a) \longrightarrow a \quad \text{(iv)}$$

$$q_\top(f(x,y)) \longrightarrow f(q_\top(x), q_\top(y)) \text{(v)} \qquad q_\top(g(x)) \longrightarrow g\left(q_{\top(x)}\right) \quad \text{(vi)}$$

$$q_\top(a) \longrightarrow a \qquad \text{(vii)}$$

Proof Show, in order, that each of the following properties holds for all trees t, with an induction on t.

$$q_\top(t) \to^* t \tag{1}$$

$$q_a(t) \to^* t \quad \Leftrightarrow \quad t = a \tag{2}$$

$$q_g(t) \to^* t \quad \Leftrightarrow \quad \exists u \in \mathcal{T}(\mathcal{F}), \quad t = g(u) \tag{3}$$

$$q_{f_2}(t) \to^* t \quad \Leftrightarrow \quad \exists u \in \mathcal{T}(\mathcal{F}), \quad t = f(a,u) \tag{4}$$

$$q_{f_1}(t) \to^* t \quad \Leftrightarrow \quad \exists u, v \in \mathcal{T}(\mathcal{F}), \quad t = f(f(a,u), g(v)) \tag{5}$$

Here is an example of such an induction proof for the property (5), assuming each of the previous ones have been proven.

Base case: Assume $t = a$. Left side of the equivalence is false, because $q_{f_1}(a)$ is blocked. Right side of the equivalence is also false, so the equivalence is true.

Case g: Assume $t = g(u)$. As before, the equivalence is true because both sides of it are false.

Case f: Assume $t = f(u, v)$.

- Assume $q_{f_1}(t) \to^* t$. The only available transition is transition (i), so we know that

$$q_{f_1}(t) \to f\Big(q_{f_2}(u), q_g(v)\Big) \to^* t$$

  In particular, we know that $q_{f_2}(u) \to^* u$ and $q_g(v) \to^* v$.

  Because property (4) holds, there exist u' such that $u = f(a, u')$. Because property (3) holds, there exist v' such that $v = g(v')$.

  In conclusion, there exist u', v' such that $t = f(f(a, u'), g(v'))$.
- Assume there exist u, v such that $t = f(f(a, u), g(v))$. In particular, we know that

$$q_{f_1}(t) \to f\Big(q_{f_2}(f(a, u)), q_g(g(v))\Big)$$

  Because property (4) holds, $q_{f_2}(f(a, u)) \to^* f(a, u)$. Because property (3) holds, $q_g(g(v)) \to^* g(v)$.

  In conclusion, $q_{f_1}(t) \to^* t$. ∎

Bottom-up DFTA: $Q = \{q_a, q_f, q_g, q_\top, q_\bot\}$, $F = \{q_\top\}$, and transitions as follow:

$$a \longrightarrow q_a \qquad f(q_a, q) \longrightarrow q_f \text{ for all } q \in Q$$

$$g(q) \longrightarrow q_g \text{ for all } q \in Q \qquad f\big(q_f, q_g\big) \longrightarrow q_\top$$

$$f(q, q') \longrightarrow q_\bot \text{ for all } (q, q') \neq (q_a, \_), \big(q_f, q_g\big)$$

Proof Show, in order, that each of the following properties holds for all trees t, with an induction on t.

$$t \to^* q_a \quad \Leftrightarrow \quad t = a$$

$$t \to^* q_g \quad \Leftrightarrow \quad \exists u \in \mathcal{T}(\mathcal{F}), \quad t = g(u)$$

$$t \to^* q_f \quad \Leftrightarrow \quad \exists u \in \mathcal{T}(\mathcal{F}), \quad t = f(a, u)$$

$$t \to^* q_\top \quad \Leftrightarrow \quad \exists u, v \in \mathcal{T}(\mathcal{F}), \quad t = f(f(a, u), g(v))$$

Exercise 2 - What is recognizable by a tree automaton?

Are the following tree languages recognizable by a bottom-up tree automaton?

1. $\mathcal{F} = \{g(1), a(0)\}$ and L the set of ground terms of even height.
2. $\mathcal{F} = \{f(2), g(1), a(0)\}$ and L the set of ground terms of even height.

Solution

1. Yes.
2. No. Remark that you can not use directly the pumping lemma to increase the size of the tree, as you can pump on a context of even size of the longest branch, and it does not create a contradiction.

   Assume that this language is recognizable by a NFTA with n states. Define:

   $$t_n = f(g^{2n+1}(a), g^{2n+2}(a))$$

   It has height $2n + 2$ and so belongs to this language. So there exists an accepting run ρ for t_n. By the pigeonhole principle, there exists $k < k'$ and a state q such that ρ goes through both $q(g^k(a))$ and $q\left(g^{k'}(a)\right)$. From that we deduce that for all $p \in \mathbb{N}$, the tree

   $$t_{n,p} = f\left(g^{2n+1+p(k'-k)}(a), g^{2n+2}(a)\right)$$

   also has an accepting run. But $t_{n,2}$ has height $2(n + k' - k) + 1$, which is odd. Contradiction. ■

Exercise 3 - Bottom-up vs Top-down

1. Recall why bottom-up NFTAs, bottom-up DFTAs and top-down NFTAs have the same expressiveness.

Solution

Bottom-up NFTAs = bottom-up DFTAS through powerset construction.

L accepted by bottom-up NFTA $\mathcal{A} = \langle Q, \mathcal{F}, G, \Delta \rangle$ iff L accepted by top-down NFTA $\mathcal{A}' = \langle Q, \mathcal{F}, G, \Delta' \rangle$, with

$$\Delta' := \{q(f) \to (q_1, ..., q_n) \mid f(q_1, ..., q_n) \to q \in \Delta\}$$

See more details in lecture.

2. Let $\mathcal{F} = \{f(2), g(1), a(0)\}$. Give a DFTA and a top-down NFTA for the set $M(t)$ of terms which have a ground instance of the term $t = f(a, g(x))$ as a subterm, i.e.

$$\{C[f(a, g(u))] \mid C \in \mathcal{C}(\mathcal{F}), u \in T(\mathcal{F})\}.$$

Solution

DFTA: $Q = \{q_a, q_g, q_\top, q_\bot\}$, $F = \{q_\top\}$ and transitions as follow:

$$a \longrightarrow q_a$$

$$g(q_\top) \longrightarrow q_\top \qquad g(q) \longrightarrow q_g \text{ with } q \neq q_\top$$

$$f(q, q') \longrightarrow q_\top \quad \text{if } (q, q') = (q_a, q_g) \text{ or } q = q_\top \text{ or } q' = q_\top$$

$$f(q, q') \longrightarrow q_\bot \quad \text{else}$$

Proof idea: Show, in order, that

1. $t \to^* q_a \iff t = a$
2. If $t \to^* q_g$ then $\exists u, t = g(u)$
3. $t \to^* q_\top \iff t \in M(t)$

Make sure that you used the transition to $q_\bot$ in your proof! It should be used in the reverse implication of 3. only.

NFTA: $Q = \{q_0, q_\top, q_a, q_g\}$, $I = \{q_0\}$ and transitions as follow:

$$q_0(f(x, y)) \longrightarrow f(q_\top(x), q_0(y)) \qquad q_0(f(x, y)) \longrightarrow f(q_0(x), q_\top(y))$$

$$q_\top(f(x, y)) \longrightarrow f(q_\top(x), f(q_\top(y))) \quad q_\top(g(x)) \longrightarrow g(q_\top(x))$$

$$q_\top(a) \longrightarrow a \qquad q_0(g(x)) \longrightarrow g(q_0(x))$$

$$q_0(f(x, y)) \longrightarrow f(q_a(x), q_g(y)) \qquad q_a(a) \longrightarrow a$$

$$q_g(g(x)) \longrightarrow g(q_\top(x))$$

Proof idea: show, in order, that

1. For all tree t, $q_\top(t) \to^* t$
2. $q_g(t) \to^* t \iff \exists u, t = g(u)$
3. $q_a(t) \to^* t \iff t = a$
4. $q_0(t) \to^* t \iff t \in M(t)$

3. Show that NFTAs and top-down DFTAs do not have the same expressiveness.

Solution

Proof by contradiction. Assume $M(t)$ can be recognized by a top-down DFTA $\mathcal{A}$. Consider $t_1 = f(t, a)$ and $t_2 = f(a, t)$. $\mathcal{A}$ must start with the same transition on both terms, say $q_0(f(x, y)) \longrightarrow f(q_L(x), q_R(y))$. Then, there exist an accepting run for $q_R(a)$ because $t_1 \in M(t)$, and conversely for $q_L(a)$. Finally, $\mathcal{A}$ accepts $f(a, a)$ which is not in $M(t)$. Contradiction.

Homework - Satisfiability

Let $\mathcal{F} = \{\mathrm{and}(2), \mathrm{or}(2), \mathrm{not}(1), 0(0), 1(0), x(0)\}$. A ground term over $\mathcal{F}$ can be viewed as a boolean formula over x.

1. Give an NFTA which recognizes the set of satisfiable boolean formulae over x. Show that it indeed does.

Let $\mathcal{F} = \left\{\mathrm{and}(2), \mathrm{or}(2), \mathrm{not}(1), 0(0), 1(0), x_1(0), ..., x_{n(0)}\right\}$, i.e. we now handle n variables instead of a single one. The same variable may appear several times in a formula, and should be evaluated consistently.

2. Give an NFTA which recognizes the set of satisfiable boolean formulae over $x_1, ..., x_n$. Show that it does.