Tree Automata and their Applications

TD n°1 : Recognizable Tree Languages and Finite Tree Automata

2021-2022

Exercise 1 : First constructions of Tree Automatas

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 :

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

Solution:

- top-down DFTA : $Q = \{q_{f,1}, q_{f,2}, q_g, q_a, q_\top\}$, $I = \{q_{f,1}\}$ and $\Delta =$
  - ⋆ $q_{f,1}(f(x, y)) \longrightarrow f(q_{f,2}(x), q_g(y))$
  - ⋆ $q_{f_2}(f(x, y)) \longrightarrow f(q_a(x), q_\top(y))$
  - ⋆ $q_g(g(x)) \longrightarrow g(q_\top(x))$
  - ⋆ $q_a(a) \longrightarrow a$
  - ⋆ $q_\top(f(x, y)) \longrightarrow f(q_\top(x), q_\top(y))$
  - ⋆ $q_\top(g(x)) \longrightarrow g(q_\top(x))$
  - ⋆ $q_\top(a) \longrightarrow a$
- DFTA : $Q = \{q_a, q_f, q_g, q_\top, q_\bot\}$, $F = \{q_\top\}$ and $\Delta =$
  - ⋆ $a \longrightarrow q_a$
  - ⋆ $f(q_a, q) \longrightarrow q_f$ for all $q \in Q$
  - ⋆ $g(q) \longrightarrow q_g$ for all $q \in Q$
  - ⋆ $f(q_f, q_g) \longrightarrow q_\top$
  - ⋆ $f(q, q') \longrightarrow q_\bot$ for all $(q, q') \neq (q_a, \_), (q_f, q_g)$

Exercise 2 : What is recognizable by an FTA ?

Are the following tree languages recognizable (by a bottom-up FTA) ?

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

Solution:

- Yes.
- No. Remark that the pumping lemma does not apply ! Assume that it 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 $k < k'$ such that $r(1.1^k) = r(1.1^{k'})$ and from that we deduce that for all $p \in \mathbb{N}$, the tree

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

  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 DTAs and top-down NFTAs have the same expressiveness.
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, ie. $M(t) = \{C[f(a, g(u))] \mid C \in \mathcal{C}(\mathcal{F}), u \in T(\mathcal{F})\}$.
3) Show that NFTAs and top-down DFTAs do not have the same expressiveness.

Solution:

- top-down NFTA : $Q = \{q_0, q_\perp, q_a, q_g\}$, $I = \{q_0\}$ and $\Delta =$
  - $\star$ $q_0(f(x, y)) \longrightarrow f(q_\perp(x), q_0(y))$
  - $\star$ $q_0(f(x, y)) \longrightarrow f(q_0(x), q_\perp(y))$
  - $\star$ $q_\perp(f(x, y)) \longrightarrow f(q_\perp(x), q_\perp(y))$
  - $\star$ $q_\perp(g(x)) \longrightarrow g(q_\perp(x))$
  - $\star$ $q_\perp(a) \longrightarrow a$
  - $\star$ $q_0(g(x)) \longrightarrow g(q_0(x))$
  - $\star$ $q_0(f(x, y)) \longrightarrow f(q_a(x), q_g(y))$
  - $\star$ $q_a(a) \longrightarrow a$
  - $\star$ $q_g(g(x)) \longrightarrow g(q_\perp(x))$
- DFTA : $Q = \{q_a, q_g, q_\top, q_\perp\}$, $F = \{q_\top\}$ and $\Delta =$
  - $\star$ $a \longrightarrow q_a$
  - $\star$ $g(q_\top) \longrightarrow q_\top$
  - $\star$ $g(q) \longrightarrow q_g$ with $q \neq q_\top$
  - $\star$ $f(q, q') \longrightarrow q_\top$ if $(q, q') = (q_a, q_g)$ or $q = q_\top$ or $q' = q_\top$
  - $\star$ $f(q, q') \longrightarrow q_\perp$ else
- Let's assume M(t) can be recognized by a top-down DFTA $\mathcal{A}$. We consider two terms $t_1 = f(t, a)$ and $t_2 = f(a, t)$. $\mathcal{A}$ must start with the same transition on both terms, let's say $q_0(f(x, y)) \longrightarrow f(q_L(x), q_R(y))$. Then, there is an accepting run for $q_R(a)$ because t_1 in M(t), and conversely for $q_L(a)$. Finally, $\mathcal{A}$ accepts $f(a, a) \notin M(t)$.

Exercise 4 : On a more abstract language.

1) Let $\mathcal{E}$ be a finite set of linear terms on $T(\mathcal{F}, \mathcal{X})$. Prove that $Red(\mathcal{E}) = \{C[t\sigma] \mid C \in \mathcal{C}(\mathcal{F}), t \in \mathcal{E}, \sigma \text{ ground substitution}\}$ is recognizable.
2) Prove that if $\mathcal{E}$ contains only ground terms, then one can construct a DFTA recognizing $Red(\mathcal{E})$ whose number of states is at most $n + 2$, where n is the number of nodes of $\mathcal{E}$.

Solution:

1) Do the case where $\mathcal{E}$ is a singleton $\{t\}$, t linear (the general case can be deduced by finite union). $Red(\{t\}$ is recognized by the following NFTA : $Q = \{q_\perp\} \cup Pos(t)$, $F = \{\epsilon\}$ and $\Delta =$
   - $\star$ $f(q_1, ..., q_n) \longrightarrow q_\perp$ for all $f \in \mathcal{F}$, q_1, ..., $q_n \in Q$
   - $\star$ $q_\perp \longrightarrow p$ for all $p \in Pos(t)$ such that $t(p)$ is a variable
   - $\star$ $f(p.1, ..., p.n) \longrightarrow p$ if $t(p) = f$
   - $\star$ $f(q_1, ..., q_n) \longrightarrow \epsilon$ for all $f \in \mathcal{F}$ and q_1, ..., $q_n \in Q$ such that there exists $i \in \{1, ..., n\}$ such that $q_i = \epsilon$

   Let $St(\mathcal{E})$ be the set of all subterms of a term in $\mathcal{E}$. Then the following DFTA $\mathcal{A}$ works : $Q = \{q_t \mid t \in St(\mathcal{E})\} \cup \{q_\perp, q_\top\}$, $F = \{q_\top\}$ and $\Delta =, \forall f \in \mathcal{F}$
2) $\star$ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_{f(t_1,...,t_n)}$ if $f(t_1, ...t_n) \in St(\mathcal{E}) \setminus \mathcal{E}$
   - $\star$ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\top$ if $f(t_1, ..., t_n) \in \mathcal{E}$
   - $\star$ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\perp$ else

- ⋆ $f(q_1, ..., q_n) \longrightarrow q_\top$ if there is at least one $q_i = q_\top$
- ⋆ $f(q_1, ..., q_n) \longrightarrow q_\bot$ else

We will, for once, and as you should at least for the first few questions of an exam, formally prove that this automaton recognizes the expected language.

We first prove by induction on the size of the terms, that $\forall t \in St(\mathcal{E}) \setminus \mathcal{E}, L(q_t) = t$.

- — If $t = a/0 \in St(\mathcal{E}) \setminus \mathcal{E}$, then, the only rule which can produce q_a is $a \longrightarrow q_a$, and we do have $L(q_a) = a$.
- — If $t = f(t_1, ..., t_n) \in St(\mathcal{E}) \setminus \mathcal{E}$, the interesting rule is then $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_{f(t_1,...,t_n)}$. Thus, $L(q_{f(t_1,...,t_n)} = \{f(x_1, \dots, x_n) | \forall 1 \leq i \leq n, x_i \in L(q_{t_i})\}$. By the induction hypothesis, we have $\forall 1 \leq i \leq n, L(q_i) = t_i$. Thus, $L(q_{f(t_1,...,t_n)} = f(t_1, ..., t_n)$.

Now, we may prove that by induction on the size n of the terms that $L(q_\bot) \supset T^{<n}(\mathcal{F}, \mathcal{X}) \setminus (Red(\mathcal{E}) \cup St(\mathcal{E})) \wedge L(q_\top) \supset Red^{<n}(\mathcal{E})$ (we denote with $L^{<n}$ all the terms of L of size at most n).

- — If $t = a/0 \notin (Red(\mathcal{E}) \cup St(\mathcal{E}))$, then we have a transition $a \longrightarrow q_\bot$.
- — If $t = a/0 \in \mathcal{E})$ then we have a transition $a \longrightarrow q_\top$.

We do have our property for $n = 0$.

- — If $t = f(t_1, ..., t_n) \in St(\mathcal{E})$, we have obtained previously that $L(q_t) = t$.
- — If $t = f(t_1, ..., t_n) \in \mathcal{E}$, the only interesting rule is $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\top$. As $f(t_1, ..., t_n) \in \mathcal{E}$, $t_1, \dots, t_n \in St(\mathcal{E})$, we obtained that $L(t_i) = q_{t_i}$, and we do have $t \in L(q_\top)$.
- — If $t = f(t_1, ..., t_n) \in Red(\mathcal{E}) \setminus \mathcal{E}$, then there exists $1 \leq i \leq n$ such that $t_i \in Red(\mathcal{E})$. Then, by induction hypothesis, $t_i \in L(q_\top)$, the others terms do reaches states are they are either in $L(q_\bot), L(q_\top)$ or some $L(q_t)$, and we can apply the transition $f(q_1, ..., q_n) \longrightarrow q_\top$, which proves that $t \in L(q_\top)$.
- — If $t = f(t_1, ..., t_n) \in T^{<n}(\mathcal{F}, \mathcal{X}) \setminus (Red(\mathcal{E}) \cup St(\mathcal{E}))$, then the only transition applicable is $f(q_1, ..., q_n) \longrightarrow q_\bot$. As by induction hypothesis the subterms are either in $L(q_\bot), L(q_\top)$ or some $L(q_t)$, we can indeed apply the transition, and we do have $t \in L(q_\bot)$.

Finally, we can conclude that $L(q_\top) = Red(\mathcal{E})$, which is the expected result.