Automates d'arbre

TD n°2 : Decision problems & tree homomorphisms

September 25, 2018

Exercise 1 : Recognizing an abstract language.

1) Let $\mathcal{E}$ be a finite set of linear terms on $T(\mathcal{F}, \mathcal{X})$. Prove that $Red(\mathcal{E}) = \big\{C[t\sigma] \mid C \in \mathcal{C}(\mathcal{F}), t \in \mathcal{E}, \sigma \text{ ground substitution}\big\}$ 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 $\Delta =$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q_\perp$ for all $f \in \mathcal{F}$, q_1, ..., $q_n \in Q$
   - ⋆ $q_\perp \longrightarrow p$ for all $p \in Pos(t)$ such that $t(p)$ is a variable
   - ⋆ $f(p.1, ..., p.n) \longrightarrow p$ if $t(p) = f$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow \epsilon$ for all $f \in \mathcal{F}$ and q_1, ..., $q_n \in Q$ such that there exists $i \in \{1, ..., n\}$ such that $q_i = \epsilon$
2) 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 $\Delta =, \forall f \in \mathcal{F}$
   - ⋆ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_{f(t_1,...,t_n)}$ if $f(t_1, ...t_n) \in St(\mathcal{E}) \setminus \mathcal{E}$
   - ⋆ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\top$ if $f(t_1, ..., t_n) \in \mathcal{E}$
   - ⋆ $f(q_{t_1}, ..., q_{t_n}) \longrightarrow q_\perp$ else
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q_\top$ if there is at least one $q_i = q_\top$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q_\perp$ 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 $L(q_\perp) \supset T^{<n}(\mathcal{F}, \mathcal{X}) \setminus (Red(\mathcal{E}) \cup St(\mathcal{E})) \wedge L(q_\top) \supset Red^{<n}(\mathcal{E})$ (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 $a \longrightarrow q_\perp$.
   - — 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 $t = f(t_1, ..., t_n) \in Red(\mathcal{E}) \backslash \mathcal{E}$, 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 $L(q_\perp), L(q_\top)$ 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 $f(q_1, ..., q_n) \longrightarrow q_\perp$. As by induction hypothesis the subterms are either in $L(q_\perp), L(q_\top)$ or some $L(q_t)$, we can indeed apply the transition, and we do have $t \in L(q_\perp)$.

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

Exercise 2 : Decisions problems

We consider the **(GII)** problem (ground instance intersection) :

Instance : t a term in $T(\mathcal{F}, \mathcal{X})$ and $\mathcal{A}$ a NFTA

Question : Is there at least one ground instance of t accepted by $\mathcal{A}$?

1) Suppose that t is linear. Prove that **(GII)** is P-complete.
2) Suppose that $\mathcal{A}$ is deterministic. Prove that **(GII)** is NP-complete.
3) Prove that **(GII)** is EXPTIME-complete.
   hint : for the hardness, reduce the intersection non-emptiness problem (admitted to be EXPTIME-complete).
4) Deduce that the complement problem :
   Instance : t a term in $T(\mathcal{F}, \mathcal{X})$ and linear terms t_1, ..., t_n
   Question : Is there a ground instance of t which is not an instance of any t_i ?
   is decidable.

Solution:

1) in P : use a construction similar to exercise 1, intersect with $\mathcal{A}$ and test the non-emptiness.
   P-hard : testing the emptiness of $\mathcal{A}$ is equivalent to testing **(GII)** on $\mathcal{A}$ and a variable.
2) in NP : guess for each variable an accessible state of $\mathcal{A}$ and verify that you can complete this to an accepting run by running the automata.
   NP-hard : We reduce **(SAT)** : let $\mathcal{F} = \{\neg(1), \vee(2), \wedge(2), \perp(0), \top(0)\}$ and $\mathcal{A}_{SAT}$ the DFTA with $Q = \{q_\top, q_\perp\}$, $F = \{q_\top\}$ and $\Delta =$
   - ⋆ $\perp \longrightarrow q_\perp$
   - ⋆ $\top \longrightarrow q_\top$
   - ⋆ $\neg(q_\alpha) \longrightarrow q_{\neg\alpha}$
   - ⋆ $\vee(q_\alpha, q_\beta) \longrightarrow q_{\alpha\vee\beta}$
   - ⋆ $\wedge(q_\alpha, q_\beta) \longrightarrow q_{\alpha\wedge\beta}$

   The language of $\mathcal{A}_{SAT}$ is the set of closed valid formulae.
   Let ϕ a CNF formula, $\phi = \bigwedge_{i=1}^{n} c_i$ where c_i are clauses. Define t_{c_i} by induction on the size of c_i :
   — if $c_i = x_j$, $t_{c_i} = x_j$
   — if $c_i = \neg x_j$, $t_{c_i} = \neg(x_j)$
   — if $c_i = x_j \vee c_i'$, $t_{c_i} = \vee(x_j, t_{c_i'})$
   — if $c_i = \neg x_j \vee c_i'$, $t_{c_i} = \vee(\neg(x_j), t_{c_i'})$
   Then $t_\phi = \wedge(t_{c_1}, \wedge(t_{c_2}, ..., \wedge(t_{c_{n-1}}, t_{c_n})...))$. ϕ is satisfiable iff a closed instance of t_ϕ is recognized by $\mathcal{A}_{SAT}$.
3) in EXP : for each coloring of t by states (exponentially many) :
   — check that the coloring of every occurrence of a variable is an accessible state (in P)
   — check that the coloring corresponds to an accepting run (in P)

— for every variable, let $\{q_1, ..., q_n\}$ be the set of the colorings of all occurrence of x. Check that $L(\mathcal{A}_{q_1}) \cap ... \cap L(\mathcal{A}_{q_n})$ is non empty where A_q is the NFTA obtained from $\mathcal{A}$ by changing the set of final states to $\{q\}$ (in P)

EXP-hard : We reduce intersection non-emptiness : let $(A_k = (Q_k, \mathcal{F}, I_k, \Delta_k))_{k\in\{1,...,n\}}$ a finite sequence of top-down NFTA (we can transform a bottom-up NFTA to a top-down one in polynomial time). We suppose that all the Q_k are disjoint. Define :

— $\mathcal{F}' = \mathcal{F} \cup \{h(n)\}$

— $t = h(x, ..., x)$

— $\tilde{\mathcal{A}} = (\bigsqcup Q_k \sqcup \{q_0\}, \mathcal{F}', \{q_0\}, \Delta' \sqcup \bigsqcup \Delta_k)$ where

$$\Delta' = \{q_0(h(x_1, ..., x_n)) \longrightarrow h(q_1(x_1), ..., q_n(x_n)) \mid \ for\ q_k \in I_k\}$$

Then $L(\mathcal{A}_1) \cap ... \cap L(\mathcal{A}_n) \neq \varnothing$ iff t has a closed instance in $L(\tilde{A})$.

4) Use question 3 and exercise 4 of TD1.

⋆⋆⋆

Homework for next week : Direct images of an homomorphism

Let $\mathcal{F} = \{f/2, g/1, a\}$ and $\mathcal{F}' = \{f'/2, g/1, a\}$. Let us consider the tree homomorphism h determined by h_F defined by : $h_{\mathcal{F}}(f) = f'(x_1, x_2), h_{\mathcal{F}}(g) = f'(x_1, x_1)$, and $h_{\mathcal{F}}(a) = a$.

1. Is $h(\mathcal{T}(\mathcal{F}))$ recognizable ?
2. Let$L_1 = \{g^i(a) | i \geq 0\}$, then L_1 is a recognizable tree language, is $h(L_1)$ recognizable ?