Tree Automata and their Applications

TD n°2 : Decision problems & tree homomorphisms

2021-2022

Exercise 1 : About Nutts

A bottom-up tree transducer (NUTT) is a tuple $U = (Q, \mathcal{F}, \mathcal{F}', Q_f, \Delta)$ where Q is a finite set (of states), $\mathcal{F}$ and $\mathcal{F}'$ are finite ranked sets (of input and output), $Q_f \subseteq Q$ (final states) and Δ is a finite set of rules of the form :

- $f(q_1(x_1), ..., q_n(x_n)) \to q(u)$ where $f \in \mathcal{F}$ and $u \in T(\mathcal{F}', \{x_1, ..., x_n\})$
- $q(x_1) \to q'(u)$ where $u \in T(\mathcal{F}', \{x_1\})$.

We say that U is linear when the right side of the rules of Δ are. This defines a rewrite system $\to_U$ on $T(\mathcal{F} \cup \mathcal{F}' \cup Q)$. The relation induced by U is then $\mathcal{R}(U) = \big\{(t, t') \mid t \in T(\mathcal{F}),\, t' \in T(\mathcal{F}'),\, t \to_U^* q(t'),\, q \in Q_f\big\}$.

1) Prove that tree morphisms are a special case of NUTT that is if $\mu : T(\mathcal{F}) \longrightarrow T(\mathcal{F}')$ is a morphism, then there exists a NUTT U_μ such that $\mathcal{R}(U_\mu) = \big\{(t, \mu(t)) \mid t \in T(\mathcal{F})\big\}$. Be sure that if μ is linear then U_μ is too.
2) Prove that the domain of a NUTT U, that is $\big\{t \in T(\mathcal{F}) \mid \exists t' \in T(\mathcal{F}'),\, (t, t') \in U\big\}$, is recognizable.
3) Prove that the image of a recognizable tree language L by a linear NUTT U, that is $\big\{t' \in T(\mathcal{F}') \mid \exists t \in L,\, (t, t') \in U\big\}$, is recognizable.

Solution:

1) $Q = \{q\}$, $Q_f = \{Q\}$ and $\Delta =$
   - ⋆ $f(q(x_1), ..., q(x_n)) \longrightarrow q(\mu(f)(x_1, ..., x_n))$ linear when μ is
2) $Q = Q_U$, $F = F_U$ and $\Delta =$
   - ⋆ $f(q_1, ..., q_n) \longrightarrow q$ if there exists u such that $f(q_1(x_1), ..., q_n(x_n)) \longrightarrow q(u) \in \Delta_U$
   - ⋆ $q \longrightarrow q'$ if there exists u such that $q(x_1) \longrightarrow q'(u) \in \Delta_U$
3) Let U a NUTT and $\mathcal{A}$ a NFTA on $\mathcal{F}$. For every pair of rules $r = f(q_1(x_1), ..., q_n(x_n)) \longrightarrow q(u) \in \Delta_U$ and $r' = f(q'_1, ..., q'_n) \longrightarrow q' \in \Delta_{\mathcal{A}}$, we define :
   - — $Q^{r,r'} = \{q_p^{r,r'} \mid p \in Pos(u)\}$
   - — $\Delta^{r,r'} =$
     - ⋆ $g(q_{p.1}^{r,r'}, ..., q_{p.k}^{r,r'}) \longrightarrow q_p^{r,r'}$ for $p \in Pos(u)$ such that $u(p) = g \in \mathcal{F}'$
     - ⋆ $(q_i, q'_i) \longrightarrow q_p^{r,r'}$ if $u(p) = x_i$ (linearity assure that we only have one of this kind for every i)
     - ⋆ $q_\epsilon^{r,r'} \longrightarrow (q, q')$

   For every rule $r = q(x) \longrightarrow q'(u) \in \Delta_U$, we define :
   - — $Q^r = \{q_p^r \mid p \in Pos(u)\} \times Q_{\mathcal{A}}$
   - — $\Delta^r =$
     - ⋆ $g((q_{p.1}^r, q''), ..., (q_{p.k}^r, q'')) \longrightarrow (q_p^r, q'')$ for $p \in Pos(u)$ such that $u(p) = g \in \mathcal{F}'$ and $q'' \in Q_{\mathcal{A}}$
     - ⋆ $(q, q'') \longrightarrow (q_p^r, q'')$ if $u(p) = x$ and $q'' \in Q_{\mathcal{A}}$ (linearity assure that we only have one of this kind)
     - ⋆ $(q_\epsilon^r, q'') \longrightarrow (q, q'')$

   Then this NFTA works :

$$\tilde{Q} = Q_U \times Q_{\mathcal{A}} \cup \bigcup_{(r,r')} Q^{r,r'} \cup \bigcup_r Q^r$$

$$\tilde{F} = F_U \times F_{\mathcal{A}}$$
$$\tilde{\Delta} = \bigcup_{(r,r')} \Delta^{r,r'} \cup \bigcup_{r} \Delta^{r}$$

Exercise 2 : Decision 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 $\mathcal{F} = \{\neg(1), \vee(2), \wedge(2), \bot(0), \top(0)\}$ and $\mathcal{A}_{SAT}$ the DFTA with $Q = \{q_\top, q_\bot\}$, $F = \{q_\top\}$ and $\Delta =$
   - ⋆ $\bot \longrightarrow q_\bot$
   - ⋆ $\top \longrightarrow q_\top$
   - ⋆ $\neg(q_\alpha) \longrightarrow q_{\neg\alpha}$
   - ⋆ $\vee(q_\alpha, q_\beta) \longrightarrow q_{\alpha\vee\beta}$
   - ⋆ $\wedge(q_\alpha, q_\beta) \longrightarrow q_{\alpha\wedge\beta}$

   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.

Exercise 3 : Path closures

Let us revisit the example from last week : $\mathcal{F} = \{f(2), g(1), a(0)\}$. Consider the set $M(t)$ of terms which have a ground instance of the term $t = f(a, g(x))$ as a subterm, i.e. $M(t) = \{C[f(a, g(u))] \mid C \in \mathcal{C}(\mathcal{F}), u \in T(\mathcal{F})\}$.

Prove that $M(t)$ is not recognizable by a finite union of languages recognizable by a top-down DFTA.

Hint : You can use without proof the following fact (prove it if you have time) : let t be a tree. The path language $\pi(t)$ is defined by :

- *if t is a constant, $\pi(t) = \{t\}$*
- *if $t = f(t_1, ..., t_n)$, $\pi(t) = \cup_{i=1}^{n}\{fiw \mid w \in \pi(t_i)\}$*

Let L be a tree language. The path language of L is $\pi(L) = \cup_{t \in L}\pi(t)$. The path closure of L is defined by

$$pathclosure(L) = \{t \mid \pi(t) \subseteq \pi(L)\}$$

L is recognizable by a top-down DFTA iff L is recognizable and path closed, i.e. $L = pathclosure(L)$.

Solution:

- non-recognizable by a top-down DFTA : it is not path-closed because if it were the case, as $f(f(a, g(a)), a)$ and $f(a, f(a, g(a)))$ are in $M(t)$, then $f(a, a)$ would be in $M(t)$ too which is absurd.
- not a finite union of languages recognizable by a top-down DFTA : assume that $M(t)$ is a finite union $\bigcup_{k=1}^{n} L_k$ where L_k is recognizable by a top-down DFTA and in particular path-closed. Let t^p be defined by induction :
  — $t^0 = f(a, t)$
  — $t^{p+1} = f(t^p, a)$
  They all belong to $M(t)$ then at least one of the L_k contains two different t^p and t^q with, say, $p < q$. As L_k is path-closed, it must also contains s^p, where s^r is defined by :
  — $s^0 = f(a, a)$
  — $s^{r+1} = f(s^r, a)$
  which does not belong to $M(t)$ (because, for example, it contains only fs and as). Contradiction.