Tree Automata and Applications: TD5
Hedge Automata

Emile Contal
http://econtal.perso.math.cnrs.fr/teaching

October 18, 2013

1 Hedge Automata

Definition 1. A nondeterministic finite hedge automaton (NFHA) over Σ is a tuple $\mathcal{A} = (\mathcal{Q}, \Sigma, \mathcal{Q}_f, \Delta)$ where $\mathcal{Q}$ is a finite set of states, $\mathcal{Q}_f \subseteq \mathcal{Q}$ is a set of final states, and Δ is a finite set of transition rules of the type $a(R) \to q$, where $R \subseteq \mathcal{Q}^*$ is a regular language over $\mathcal{Q}$. These languages R occurring in the transition rules are called horizontal languages.

A run of $\mathcal{A}$ on a tree $t \in T(\Sigma)$ is a tree $r \in T(\mathcal{Q})$ with the same domain as t such that for each node $p \in \mathsf{Pos}(r)$ with $a = t(p)$ and $q = r(p)$ there is a transition rule $a(R) \to q$ of $\mathcal{A}$ with $r(p1) \cdot r(pn) \in R$, where n denotes the number of successors of p. In particular, to apply a rule at a leaf, the empty word ϵ has to be in the horizontal language of the rule.

An unranked tree t is accepted by $\mathcal{A}$ if there is a run r of $\mathcal{A}$ on t whose root is labeled by a final state, i.e. with $r(\epsilon) \in \mathcal{Q}_f$. The language $\mathcal{L}(\mathcal{A})$ of $\mathcal{A}$ is the set of all unranked trees accepted by $\mathcal{A}$.

Exercise 1. Let $\Sigma = \{or, and, not, 0, 1\}$. Define a NFHA $\mathcal{A}$ such that $\mathcal{L}(\mathcal{A})$ is the set of all trees that form correct Boolean expressions evaluating to true.

Exercise 2. Let $\Sigma = \{a, b, c\}$. Define a NFHA $\mathcal{A}$ such that $\mathcal{L}(\mathcal{A})$ is the set of all trees such that there exists two nodes labelled b whose greatest common ancestor is labeled c.

2 First-Child-Next-Sibling Encoding

Definition 2. We define the First-Child-Next-Sibling encoding of an unranked tree on the alphabet Σ into a tree of rank 2 on the alphabet $\mathcal{F}^{\Sigma}_{\mathsf{FCNS}}$ with the following function fcns:

- $\mathcal{F}^{\Sigma}_{\mathsf{FCNS}} = \big\{a(,) \mid a \in \Sigma\big\} \cup \{\#\}$
- for all $a \in \Sigma$, $\mathsf{fcns}(a) = a(\#, \#)$
- for a tree $t = a(t_1 \cdots t_n)$, $\mathsf{fcns}(t) = a\big(\mathsf{fcns}(t_1 \cdots t_n), \#\big)$
- for a hedge $h = t_1 \cdots t_n$ with $n \geq 2$, $\mathsf{fcns}(h) = \mathsf{fcns}(t_1)\big[\mathsf{fcns}(t_2 \cdots t_n)\big]_2$, where $t[t']_p$ denotes the tree obtained from t by replacing the subtree at position p by t'.

We extend fcns to the sets of trees in the usual way:

$$\mathsf{fcns}(L) = \{\mathsf{fcns}(t) \mid t \in L\} \ .$$

Example exercise. Give the FCNS encoding for the unranked tree $a\big(c(b)cd(bb)\big)$.

Exercise 3. Prove that if $L \subseteq T(\Sigma)$ is hedge recognizable, then $\mathsf{fcns}(L)$ is recognizable.

3 Extension Encoding and Deterministic Step-wise Hedge Automata

Definition 3. We define the extension encoding of an unranked tree on the alphabet Σ into a tree of rank 2 on the alphabet $\mathcal{F}^{\Sigma}_{\mathsf{EXT}}$ with the following function ext:

- $\mathcal{F}^{\Sigma}_{\mathsf{EXT}} = \{@(,)\} \cup \{a \mid a \in \Sigma\}$
- for $a \in \Sigma$, $\mathsf{ext}(a) = a$
- for a tree $t = a(t_1 \cdots t_n)$ with $n \geq 1$, $\mathsf{ext}(t) = @\Big(\mathsf{ext}\big(a(t_1 \cdots t_{n-1})\big), \mathsf{ext}(t_n)\Big)$.

We extend ext to the sets of trees in the usual way:

$$\mathsf{ext}(L) = \{\mathsf{ext}(t) \mid t \in L\} \ .$$

Example exercise. Give the EXT encoding for the unranked tree $a\big(c(b)cd(bb)\big)$.

Definition 4. A Deterministic Step-wise Hedge Automaton (DSHA) is a tuple $\mathcal{A} = (\mathcal{Q}, \Sigma, \delta_0, \mathcal{Q}_f, \delta)$, where $\mathcal{Q}$, Σ, and $\mathcal{Q}_f$ are as usual, $\delta_0 : \Sigma \to \mathcal{Q}$ is a function assigning to each letter of the alphabet an initial state, and $\delta : \mathcal{Q} \times \mathcal{Q} \to \mathcal{Q}$ is the transition function. To define how such an automaton works on trees we first define how it reads sequences of its own states. For $a \in \Sigma$ let $\delta_a : \mathcal{Q}^* \to \mathcal{Q}$ be defined inductively by $\delta_a(\epsilon) = \delta_0(a)$, and $\delta_a(wq) = \delta(\delta_a(w), q)$. This corresponds to the view of $\mathcal{A}$ as a word automaton reading its own states.

For a tree t and a state q of $\mathcal{A}$ we define the relation $t \xrightarrow[\mathcal{A}]{*} q$ as follow. Let $t = a(t_1 \cdots t_n)$ and $t_i \xrightarrow[\mathcal{A}]{*} q_i$ for each $i \in \{1, \dots, n\}$. Then $t \xrightarrow[\mathcal{A}]{*} q$ for $q = \delta_a(q_1 \cdots q_n)$. For $n = 0$, this means $q = \delta_0(a)$.

Exercise 4. Let $\mathcal{A} = (\mathcal{Q}, \Sigma, \delta_0, \mathcal{Q}_f, \delta)$ be a DSHA and $t, t' \in T(\Sigma)$ with $t \xrightarrow[\mathcal{A}]{*} q$ and $t' \xrightarrow[\mathcal{A}]{*} q'$ for $q, q' \in \mathcal{Q}$. Prove that $t@t' \xrightarrow[\mathcal{A}]{*} \delta(q, q')$.

Exercise 5. Let $\mathcal{A}$ be a DSHA. Prove that $\mathsf{ext}\big(\mathcal{L}(\mathcal{A})\big) = \mathcal{L}\big(\mathsf{ext}(\mathcal{A})\big)$.

Exercise 6. Prove that if $L \subseteq T(\Sigma)$ is hedge recognizable, then $\mathsf{ext}(L)$ is recognizable.

Exercise 7. Prove that the class of recognizable unranked tree languages is closed under union, under intersection, and under complementation.