Shape Preserving Bottom-Up Tree Transducers

Zsolt Gazdag

A generalized sequential machine (gsm) is a system $M = (Q, \Sigma, \Delta, q_0, \delta, F)$, where Q is the set of states; Σ and Δ are the input and the output alphabets, respectively; q_0 is the initial state; $F \subseteq Q$ is the set of final states; and δ, the transition function, is a mapping from $Q \times \Sigma$ to the finite subsets of $Q \times \Delta^*$. Then δ extends from $Q \times \Sigma^*$ to the finite subsets of $Q \times \Delta^*$ in a standard way and the translation defined by M is the set $\tau_M = \{(x, y) \in \Sigma^* \times \Delta^* \mid (q, y) \in \delta(q_0, x)$ for some $q \in F\}$.

In general the length of an input string $x \in \Sigma^*$ and of an output string $y \in \tau_M(x)$ is not the same, however if τ_M has this property then M is called a length preserving gsm. For instance if M is a Mealy automaton, i.e., δ maps to the subsets of $Q \times \Delta$, then M is length preserving. It is a well known result that a gsm M is length preserving if and only if it is equivalent to a Mealy automaton [1], [3].

In [2] this result was generalized to top-down tree transducers and it remained an open question whether this result can be generalized to bottom-up tree transducers. While a gsm operates over strings, a bottom-up tree transducer works on terms (or rather trees), which are called also trees.

More exactly, a bottom-up tree transducer is a system $M = (Q, \Sigma, \Delta, q_0, R)$, where Q is the set of states; Σ and Δ are the input and the output ranked alphabets, respectively, and q_0 is the final state. Moreover, R is a finite set of (rewriting) rules of the form $\sigma(q_1(x_1), \dots, q_k(x_k)) \to q(r)$, where $q, q_1, \dots, q_k \in Q$, σ is an input symbol of arity k from Σ, and r is a term over Δ which may contain also variables from the set $\{x_i | 1 \le i \le k\}$. Using the rewriting rules, an input tree s over Σ, can be rewritten to a term of the form $q_0(t)$ where t is an output tree over Δ. We denote this fact by $s \Rightarrow_M^* q_0(t)$. Now the tree transformation induced by M is the set $\tau_M = \{(s, t) \in T_\Sigma \times T_\Delta \mid s \Rightarrow_M^* q_0(t)\}$, where T_Σ and T_Δ denote the set of trees over Σ and Δ, respectively.

Since trees generalize strings, more or less it should be clear that tree transducers generalize gsm's. Two trees $s \in T_\Sigma$ and $t \in T_\Delta$, have the same shape if the domains of s and t are the same, i.e., they differ only in the labels of their nodes. One can also find out easily that a natural generalization of the length preserving property of gsm's for tree transducers is the shape preserving property. A bottom-up tree transducer M is shape preserving if for every input tree $s \in T_\Sigma$ and output tree $t \in \tau_M(s)$, s and t have the same shape.

As the main result of this paper we show that every shape preserving bottom-up tree transducer is equivalent to a bottom-up relabeling tree transducer, i.e., a bottom-up tree transducer of which the rules have the form $\sigma(q_1(x_1), \dots, q_k(x_k)) \to q(\delta(x_1, \dots, x_k))$, where δ is an output symbol of arity k from Δ. This result naturally generalizes the corresponding one for gsm's.

References

[1] S. Eilenberg. Automata, Languages and Machines, Vol. A, Academic Press, 1974.

[2] Z. Fülöp and Zs. Gazdag. Shape preserving top-down tree transducers. Theoretical Computer Science, 304:315–339, 2003.

[3] J. Leguy. Transductions rationelles décroissantes. R. A. I. R. O. Theoretical Informatics, 5:141–148, 1981.