Characterizing Families of Tree Languages by Syntactic Monoids

SAEED SALEHI

http://staff.cs.utu.fi/staff/saeed/

Szeged — 2006

Trees as Terms

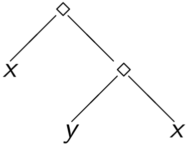
Ranked alphabet Σ, Leaf alphabet X
Σ_0 constants / Σ_m m-ary functions
$T(\Sigma, X) =$ set of trees with
node labels from Σ / leaf labels from $\Sigma_0 \cup X$

$\mathrm{T}(\Sigma, X)$ is the smallest set satisfying

- $\Sigma_0 \cup X \subseteq \mathrm{T}(\Sigma, X)$, and
- $t_1, \ldots, t_m \in \mathrm{T}(\Sigma, X) \,\&\, f \in \Sigma_m \Rightarrow f(t_1, \cdots, t_m) \in \mathrm{T}(\Sigma, X)$.

Example

$$\Sigma^S = \{\diamond/2\},\ X = \{x, y\}$$



$$= \diamond\big(x, \diamond(y, x)\big) \in T(\Sigma^S, X)$$

$$x \diamond (y \diamond x)$$

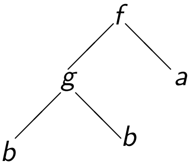
Example (Words as Trees)

$$\Lambda = \Lambda_1 = \{a/1, b/1, \ldots\},\ Y = \{\epsilon\}$$

$$\begin{array}{c} a \\ | \\ a \\ | \\ b \\ | \\ \epsilon \end{array} \quad = baa = a(a(b(\epsilon))) \in T(\Lambda, Y)$$

Example (Ground Trees)

$$\Gamma = \Gamma_2 \cup \Gamma_0: \quad \Gamma_2 = \{f, g\},\ \Gamma_0 = \{a, b\}$$

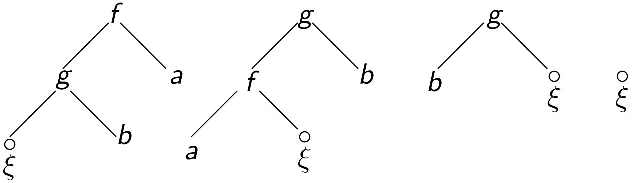


$$= \ f\big(g(b,b), a\big) \in \mathrm{T}(\Gamma, \emptyset) = \mathrm{T}_\Gamma$$

Contexts

Contexts $C(\Sigma, X)$: $(\Sigma, X \cup \{\xi\})$-trees in which the new special symbol ξ appears exactly once.

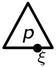
Examples: $\Gamma = \Gamma_2 \cup \Gamma_0$: $\Gamma_2 = \{f, g\}$, $\Gamma_0 = \{a, b\}$



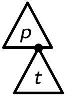
$\in \mathrm{C}(\Gamma, \emptyset) = \mathrm{C}_\Gamma$

Trees and Contexts

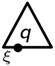
For context p and term or context s,
$p[s]$ results from p by putting s in place of ξ.

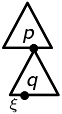
Write $p =$ 

. If 

is a tree, then $p[t] =$ 

is a tree also,

and if $q =$ 

is another context, then $p[q] =$ 

is a context as well.

$\Big\langle C(\Sigma, X), \circ \Big\rangle$ is a monoid with $p \circ q = p[q]$

Tree Languages

Any $T \subseteq \mathrm{T}(\Sigma, X)$ is a ΣX-tree language.

Two trees $t, s \in \mathrm{T}(\Sigma, X)$ are congruent w.r.t T (synonymous in the language T) iff they appear in the same context (in T):
$t \sim^T s \iff \forall P \in \mathrm{C}(\Sigma, X) \, \{\!\{ P[t] \in T \leftrightarrow P[s] \in T \}\!\}$.

Also for contexts $P, Q \in \mathrm{C}(\Sigma, X)$, monoid T-congruence is
$P \approx^T Q \iff$
$\forall R \in \mathrm{C}(\Sigma, X) \, \forall t \in \mathrm{T}(\Sigma, X) \, \{\!\{ R[P[t]] \in T \leftrightarrow R[Q[t]] \in T \}\!\}$.

The syntactic monoid $\mathrm{SM}(T)$ of T is the monoid $\mathrm{C}(\Sigma, X)/\!\approx^T$.
The tree language T is recognizable (regular) iff $\mathrm{SM}(T)$ is finite.

Example

$$\Gamma = \Gamma_2 \cup \Gamma_0: \quad \Gamma_2 = \{f, g\},\ \Gamma_0 = \{a, b\}$$

$T_1 = \{t \in \mathrm{T}_\Gamma \mid \mathtt{root}(t) = f\}$ (1-Definite tree langauge)
$\mathrm{SM}(T_1) = \{\mathfrak{f}, \mathfrak{g}, \mathfrak{1}\}$: $\mathfrak{1} =$ *identity*, $\mathfrak{f} \circ \mathfrak{f} = \mathfrak{f} \circ \mathfrak{g} = \mathfrak{f}$, $\mathfrak{g} \circ \mathfrak{f} = \mathfrak{g} \circ \mathfrak{g} = \mathfrak{g}$.
$\mathfrak{f} = \{$contexts with root $f\}$; $\mathfrak{g} = \{$contexts with root $g\}$; $\mathfrak{1} = \{\xi\}$.

$T_2 = \{t \in \mathrm{T}_\Gamma \mid \text{left-most leaf}\,(t) = a\}$ (non-definite)
$\mathrm{SM}(T_2) = \{\mathfrak{a}, \mathfrak{b}, \mathfrak{1}\}$: $\mathfrak{1} =$ *identity*, $\mathfrak{a} \circ \mathfrak{b} = \mathfrak{a} \circ \mathfrak{a} = \mathfrak{a}$, $\mathfrak{b} \circ \mathfrak{a} = \mathfrak{b} \circ \mathfrak{b} = \mathfrak{b}$.
$\mathfrak{a} = \{$contexts with left-most leaf $a\}$;
$\mathfrak{b} = \{$contexts with left-most leaf $b\}$;
$\mathfrak{1} = \{$contexts with left-most leaf $\xi\}$.

Left-most leaf $\mathtt{left}(t)$:
- $\mathtt{left}(c) = c, \quad c \in \Sigma_0 \cup X$;
- $\mathtt{left}(f(t_1, \ldots, t_m)) = \mathtt{left}(t_1)$.

$\mathrm{SM}(T_1) \cong \mathrm{SM}(T_2)$ are isomorphic !

Families of Tree Languages

For a fixed Σ, mapping $X \mapsto \mathscr{V}(X)$
$\mathscr{V} = \{\mathscr{V}(X)\}$, $\mathscr{V}(X)$ is a set of ΣX-tree languages for each X.

Generalized families of tree languages

$\mathscr{W} = \{\mathscr{W}(\Sigma, X)\}$, where $\mathscr{W}(\Sigma, X)$ is a set of ΣX-tree languages for each pair $\langle \Sigma, X \rangle$.

By considering syntactic monoids we loose track of the ranked alphabets; so generalized families of tree languages are what can be defined by varieties of monoids:
Variety of Finite Monoids $\mathbf{M} \mapsto \{\mathbf{M}^t(\Sigma, X)\}$

$$\mathbf{M}^t(\Sigma, X) = \{T \subseteq \mathrm{T}(\Sigma, X) \mid \mathrm{SM}(T) \in \mathbf{M}\}.$$

Varieties of Tree Languages

A family $\{\mathscr{V}(X)\}$ of tree languages is a variety if for any $T, T' \in \mathscr{V}(X)$

- $T \cap T', T \cup T', T^{\complement} \in \mathscr{V}(X)$;
- for $P \in \mathrm{C}(\Sigma, X)$,
  $P^{-1}(T) = \{t \in \mathrm{T}(\Sigma, X) \mid P[t] \in T\} \in \mathscr{V}(X)$;
- for morphism $\varphi : \mathrm{T}(\Sigma, Y) \to \mathrm{T}(\Sigma, X)$,
  $T\varphi^{-1} = \{t \in \mathrm{T}(\Sigma, Y) \mid t\varphi \in T\} \in \mathscr{V}(Y)$.

A morphism $\varphi : \mathrm{T}(\Sigma, Y) \to \mathrm{T}(\Sigma, X)$ maps
- any $y \in Y$ to arbitrary $y\varphi \in \mathrm{T}(\Sigma, X)$,
- $c \in \Sigma_0$ to $c\varphi = c$, and
- $f(t_1, \cdots, t_m)\varphi = f(t_1\varphi, \cdots, t_m\varphi)$.

Varieties of Finite Monoids

$M \preccurlyeq N$: M is a sub-monoid of a quotient of N

Variety of finite monoids $\mathbf{M}$: if $M_1, \ldots, M_n \in \mathbf{M}$ and $M \preccurlyeq M_1 \times \cdots \times M_n$, then $M \in \mathbf{M}$.

- $\mathrm{SM}(T \cap T'), \mathrm{SM}(T \cup T') \preccurlyeq \mathrm{SM}(T) \times \mathrm{SM}(T')$;
- $\mathrm{SM}(T^{\complement}) \cong \mathrm{SM}(T)$;
- $\mathrm{SM}(P^{-1}(T)), \mathrm{SM}(T\varphi^{-1}) \preccurlyeq \mathrm{SM}(T)$.

Tree Homomorphisms

Tree Homomorphism $\varphi : \mathrm{T}(\Omega, Y) \to \mathrm{T}(\Sigma, X)$

new variables $\xi_1, \xi_2, \ldots$.
- $\varphi_Y : Y \to \mathrm{T}(\Sigma, X)$
- $\varphi_m : \Omega_m \to \mathrm{T}(\Sigma, X \cup \{\xi_1, \ldots, \xi_m\})$ $(m \geq 0)$

- $y\varphi = y\varphi_Y$;
- $c\varphi = \varphi_0(c)$;
- $f(t_1, \ldots, t_m)\varphi = \varphi_m(f)[\![\, \xi_1 \leftarrow t_1\varphi, \ldots, \xi_m \leftarrow t_m\varphi \,]\!]$.

Regular Tree Homomorphism:
each ξ_i appears exactly once in $\varphi_m(f)$ for each $m \geq 0$, $f \in \Omega_m$.

Example

$$\Gamma = \Gamma_2 \cup \Gamma_0: \quad \Gamma_2 = \{f, g\},\ \Gamma_0 = \{a, b\}$$

Define $\psi : \mathrm{T}_\Gamma \to \mathrm{T}_\Gamma$ by

- $\psi_2(f) = f(a, f(\xi_1, \xi_2)), \quad \psi_2(g) = g(b, g(\xi_1, \xi_2))$;
- $\psi_0(a) = g(b, b), \quad \psi_0(b) = b$.

ψ is a regular tree homomorphism; e.g.

$g(b, b)\psi = g(b, g(b, b))$;

$f(g(b, b), a)\psi = f(a, f(g(b, g(b, b)), g(b, b)))$.

Also, $T_2\psi^{-1} = T_1$. $\quad \left[\texttt{left}\,(t\psi) = a \iff \texttt{root}(t) = f\right]$.

Regular Tree Homomorphisms

$\varphi : \mathrm{T}(\Omega, Y) \to \mathrm{T}(\Sigma, X)$ can be extended to contexts
$\varphi_* : \mathrm{C}(\Omega, Y) \to \mathrm{C}(\Sigma, X)$ by putting $\varphi_*(\xi) = \xi$.

In the above example:
$g(b, \xi)\psi_* = g(b, g(b, \xi))$;
$f(a, \xi)\psi = f(a, f(g(b, b), \xi))$;
$g(f(a, \xi), b)\psi_* = g(b, g(f(a, f(g(b, b), \xi)), b))$.

Regular Tree Homomorphisms and Syntactic Monoids

$\varphi : \mathrm{T}(\Omega, Y) \to \mathrm{T}(\Sigma, X) \quad \varphi_* : \mathrm{C}(\Omega, Y) \to \mathrm{C}(\Sigma, X)$
is full with respect to $T \subseteq \mathrm{T}(\Sigma, X)$ if
for any $t \in \mathrm{T}(\Sigma, X)$ and $P \in \mathrm{C}(\Sigma, X)$ there are
$s \in \mathrm{T}(\Omega, Y)$ and $Q \in \mathrm{C}(\Omega, Y)$ such that
$s\varphi \sim^T t$ and $Q\varphi_* \approx^T P$.

In other words, φ and φ_* are surjective up to T.

For any such $\varphi : \mathrm{T}(\Omega, Y) \to \mathrm{T}(\Sigma, X)$ and $T \subseteq \mathrm{T}(\Sigma, X)$

- $\mathrm{SM}(T\varphi^{-1}) \preccurlyeq \mathrm{SM}(T)$.
- If φ is full w.r.t T, then $\mathrm{SM}(T\varphi^{-1}) \cong \mathrm{SM}(T)$.

A Variety Theorem for Monoids

A generalized family $\mathscr{W} = \{\mathscr{W}(\Sigma, X)\}$ is M-variety if for any $T, T' \in \mathscr{W}(\Sigma, X)$

- $T \cap T', T \cup T', T^{\complement} \in \mathscr{W}(\Sigma, X)$;
- for any $P \in \mathrm{C}(\Sigma, X)$, $P^{-1}(T) \in \mathscr{W}(\Sigma, X)$;
- for any regular tree homomorphism $\varphi : \mathrm{T}(\Omega, Y) \to \mathrm{T}(\Sigma, X)$, $T\varphi^{-1} \in \mathscr{W}(\Omega, Y)$;
- for any regular tree homomorphism $\varphi : \mathrm{T}(\Omega, Y) \to \mathrm{T}(\Sigma, X)$ full with respect to $U \subseteq \mathrm{T}(\Sigma, X)$, if $U\varphi^{-1} \in \mathscr{W}(\Omega, Y)$ then $U \in \mathscr{W}(\Sigma, X)$;
- for any unary $\Lambda = \Lambda_1$, if $Y \subseteq Y'$ then $\mathscr{W}(\Lambda, Y) \subseteq \mathscr{W}(\Lambda, Y')$.

A Variety Theorem for Monoids

For any variety of finite monoids $\mathbf{M}$, the family $\mathbf{M}^t = \{\mathbf{M}^t(\Sigma, X)\}$ where $\mathbf{M}^t(\Sigma, X) = \{T \subseteq \mathrm{T}(\Sigma, X) \mid \mathrm{SM}(T) \in \mathbf{M}\}$ is an M-variety;

and conversely, any M-variety $\mathscr{W}$ is definable by monoids, i.e., there is a variety of finite monoids $\mathbf{M}$ such that $\mathscr{W} = \mathbf{M}^t$.

Example

Semilattice Monoids: commutative and idempotent;

$$\alpha, \beta \in \langle M, \centerdot \rangle : \ \alpha \centerdot \beta = \beta \centerdot \alpha \quad \& \quad \alpha \centerdot \alpha = \alpha.$$

Semilattice Tree Languages: $T \subseteq \mathrm{T}(\Sigma, X) \ni t, t'$

$t \in T \ \& \ c(t) = c(t') \ \Rightarrow \ t' \in T$;

$c(t) = \{$set of symbols from $\Sigma \cup X$ appearing in $t\}$.

[Unions of $\{\mathrm{T}(\Sigma', X')\}_{\Sigma' \subseteq \Sigma, X' \subseteq X}$]

(non-)Example

1-Definite tree languages are finite unions of languages of the form $\{t \mid \mathrm{root}(t) = f\}$ for an $f \in \Sigma \cup X$.
$\big($ If $f \in \Sigma_0 \cup X$ then $\{t \mid \mathrm{root}(t) = f\} = \{f\}$. $\big)$

The family Def_1 of 1-definite tree languages is a generalized variety of tree languages, not definable by monoids (nor by semigroups).

(non-)Example

In our example we have $T_2\psi^{-1} = T_1 \in \mathrm{Def}_1(\Gamma, \emptyset)$ and ψ is a regular tree homomorphism full w.r.t T_2, but $T_2 \notin \mathrm{Def}_1(\Gamma, \emptyset)$.

$a \sim^{T_2} f(b,b)\psi$; $\quad b \sim^{T_2} b\psi$;
$\mathfrak{a} \approx^{T_2} f(b,\xi)\psi_*$; $\quad \mathfrak{b} \approx^{T_2} g(b,\xi)\psi_*$; $\quad \mathfrak{1} \approx^{T_2} \xi\psi_*$.

Indeed T_2 is not a definite tree language;
but $\mathrm{SM}(T_2) \cong \mathrm{SM}(T_1)$ for a definite T_1.

This refutes a statement claimed in 1989.

❀ Thank You ! ❀

Saeed Salehi, Varieties of tree languages definable by syntactic monoids, *Acta Cybernetica* **17** (2005), 21–41.

❁ ❁ ❁ ❁ ❁ ❁ ❁ ❁ ❁ ❁

Tatjana Petković & Saeed Salehi, Positive varieties of tree languages, *Theoretical Computer Science* **347** (2005), 1–35.
