TECHNISCHE UNIVERSITÄT DRESDEN

Deutsche Forschungsgemeinschaft DFG

Freunde und Förderer der Informatik der TU Dresden e.V.

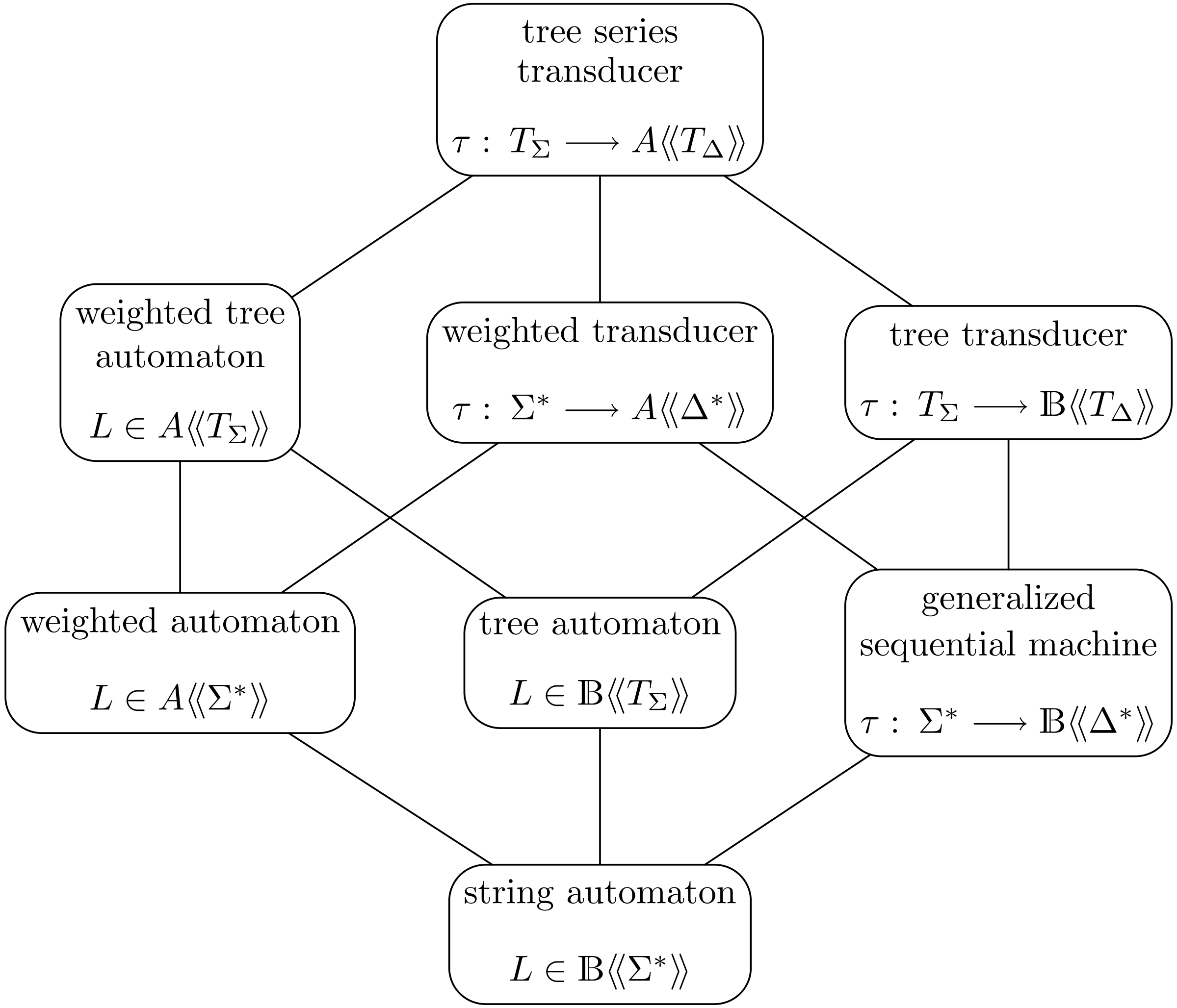
Relating Tree Series Transducers and Weighted Tree Automata

Andreas Maletti

December 17, 2004

1. Motivation and Introductory Example
2. Semirings and DM-Monoids
3. Bottom-Up DM-Monoid Weighted Tree Automata
4. Establishing a Relationship

Generalisation Hierarchy



Known Relations and Problems

- *String-based:*
  Theorem: Every gsm-mapping can be computed by a weighted automaton.
  Proof Idea: Extend monoid $(\Delta^*, \circ, \varepsilon)$ to semiring $(\mathcal{P}(\Delta^*), \cup, \circ, \emptyset, \{\varepsilon\})$

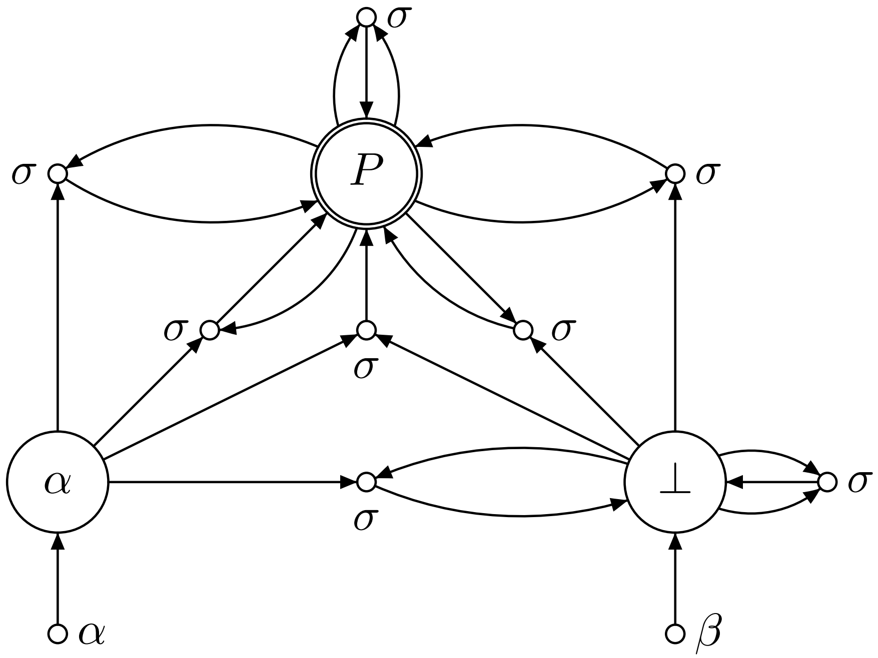
  Theorem: Weighted transductions can be computed by weighted automata.

- *Tree-based:*
  Problem: Are tree transductions computable by weighted tree automata ?

  Problem: Are tree series transformations computable by weighted tree automata ?

Tree Pattern Matching

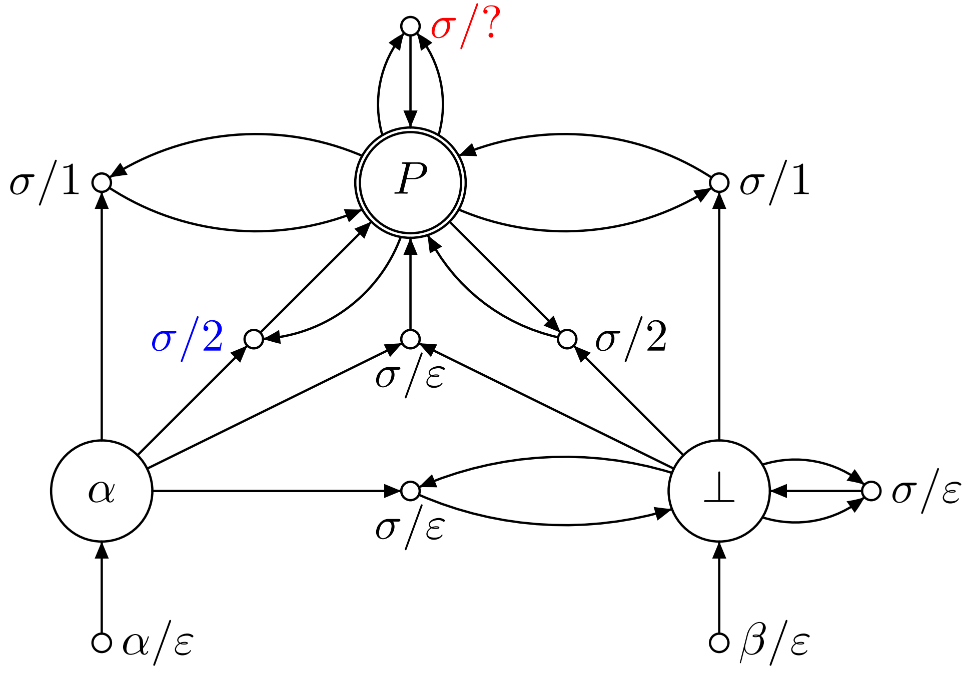
A deterministic (bottom-up) tree automaton matching the pattern $\sigma(\alpha, x)$



If pattern found, accepts tree. Otherwise reject.

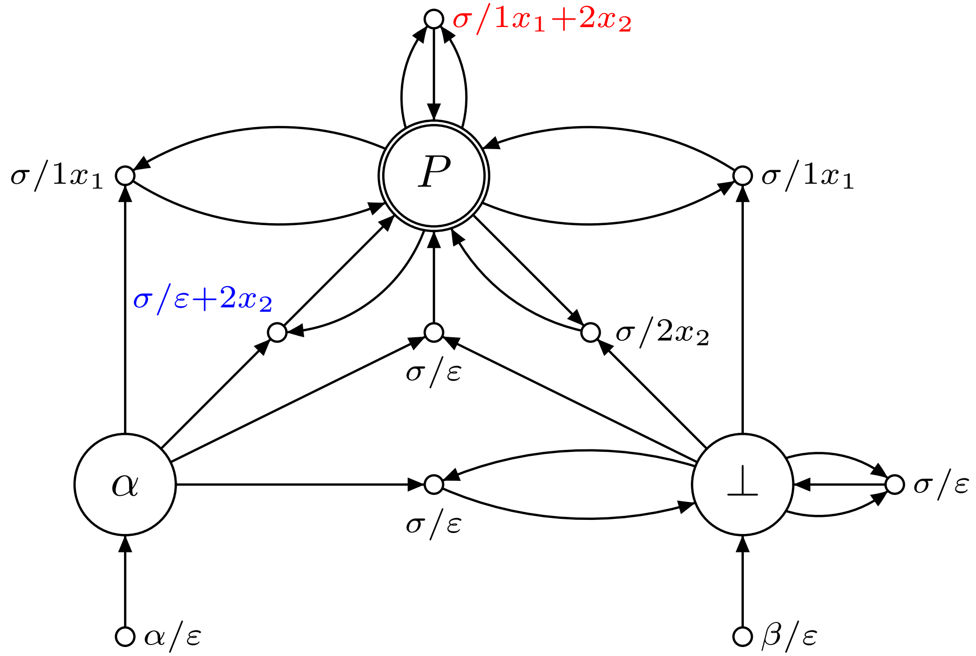
Extended Tree Pattern Matching

Towards a deterministic (bottom-up) weighted tree automaton computing an occurrence of pattern $\sigma(\alpha, x)$



Extended Tree Pattern Matching

A deterministic tree transducer computing the occurrences of pattern $\sigma(\alpha, x)$



Computes the set of occurrences of $\sigma(\alpha, x)$ in input tree.

Complete Monoids

- $\mathcal{A} = (A, \bigoplus)$ complete monoid, iff

(C1) $\bigoplus_{i \in \{j\}} a_i = a_j$,

(C2) $\bigoplus_{j \in J} (\bigoplus_{i \in I_j} a_i) = \bigoplus_{i \in I} a_i$, if $I = \bigcup_{j \in J} I_j$ is a partition.

- $\mathcal{A}$ naturally ordered, iff $\sqsubseteq$ is partial order

$$a \sqsubseteq b \quad \Longleftrightarrow \quad (\exists c \in A) : \ a \oplus c = b$$

- $\mathcal{A}$ continuous, iff $\mathcal{A}$ *naturally ordered* and *complete* and

$$\bigoplus_{i \in I} a_i \sqsubseteq a \Longleftrightarrow \bigoplus_{i \in E} a_i \sqsubseteq a \text{ for all finite } E \subseteq I$$

Semirings

- $(A, \oplus, \odot, \mathbf{0}, \mathbf{1})$ semiring, iff

  (i) $(A, \oplus, \mathbf{0})$ commutative monoid,

  (ii) $(A, \odot, \mathbf{1})$ monoid,

  (iii) $\mathbf{0}$ absorbing element with respect to $\odot$, and

  (iv) $\odot$ (left and right) distributes over $\oplus$.

- $(A, \odot, \mathbf{0}, \mathbf{1}, \bigoplus)$ complete semiring, iff

(S1) $(A, \oplus, \odot, \mathbf{0}, \mathbf{1})$ semiring,

(S2) $(A, \bigoplus)$ complete monoid, and

(S3) $a \odot (\bigoplus_{i \in I} a_i) = \bigoplus_{i \in I} (a \odot a_i)$ and $(\bigoplus_{i \in I} a_i) \odot a = \bigoplus_{i \in I} (a_i \odot a)$.

Examples of Semirings

- complete natural numbers semiring $\mathbb{N}_\infty = (\mathbb{N} \cup \{\infty\}, +, \cdot, 0, 1)$,
- tropical semiring $\mathrm{Trop} = (\mathbb{N} \cup \{\infty\}, \min, +, \infty, 0)$,
- Boolean semiring $\mathbb{B} = (\{\bot, \top\}, \vee, \wedge, \bot, \top)$,
- formal language semiring $\mathrm{Lang}_\Sigma = (\mathcal{P}(\Sigma^*), \cup, \circ, \emptyset, \{\varepsilon\})$

Semiring	commutative	complete	naturally ordered	continuous
$\mathbb{N}_\infty$	yes	yes	yes	yes
Trop	yes	yes	yes	yes
$\mathbb{B}$	yes	yes	yes	yes
Lang_Σ	NO	yes	yes	yes

Excursion: Tree Series

$(A, \bigoplus)$ complete monoid, Σ ranked alphabet, and $X_k = \{x_1, \ldots, x_k\}$.

- Tree series is mapping $\psi : T_\Sigma(X_k) \longrightarrow A$
- $A\langle\langle T_\Sigma(X_k)\rangle\rangle$ set of all tree series
- Sum $(\bigoplus_{i \in I} \psi_i, t) = \bigoplus_{i \in I} (\psi_i, t)$
- $(A\langle\langle T_\Sigma(X_k)\rangle\rangle, \bigoplus)$ complete monoid

$(A, \odot, \mathbf{0}, \mathbf{1}, \bigoplus)$ complete semiring

- Tree series substiution of $\psi_1, \ldots, \psi_k \in A\langle\langle T_\Sigma\rangle\rangle$ into $\psi \in A\langle\langle T_\Sigma(X_k)\rangle\rangle$ is

$$\psi \longleftarrow (\psi_1, \ldots, \psi_k) = \bigoplus_{\substack{t \in T_\Sigma(X_k), \\ (\forall i \in [k]): t_i \in T_\Sigma}} \Big((\psi, t) \odot \bigodot_{i \in [k]} (\psi_i, t_i) \Big) \, t[t_1, \ldots, t_k]$$

Complete DM-Monoids

$(D, \sum)$ complete monoid, Ω ranked set

- $(D, \Omega, \sum)$ distributive multi-operator monoid (DM-monoid), iff

$$\omega(\sum_{i_1 \in I_1} d_{i_1}, \ldots, \sum_{i_k \in I_k} d_{i_k}) = \sum_{(\forall j \in [k]) \colon i_j \in I_j} \omega(d_{i_1}, \ldots, d_{i_k}).$$

Examples:

- $(A, \odot, \bigoplus)$ complete semiring, $\Omega_{(k)} = \{ \underline{a}_{(k)} \mid a \in A \}$ with

$$\underline{a}_{(k)}(d_1, \ldots, d_k) = a \odot d_1 \odot \cdots \odot d_k$$

  Then $(A, \Omega, \bigoplus)$ complete DM-monoid

- $(A, \odot, \mathbf{0}, \mathbf{1}, \bigoplus)$ complete semiring, $\Omega_{(k)} = \{ \underline{\psi}_{(k)} \mid \psi \in A\langle\langle T_\Delta(X_k) \rangle\rangle \}$ with

$$\underline{\psi}_{(k)}(\psi_1, \ldots, \psi_k) = \psi \longleftarrow (\psi_1, \ldots, \psi_k)$$

  Then $(A\langle\langle T_\Delta \rangle\rangle, \Omega, \bigoplus)$ complete DM-monoid

DM-Monoid Weighted Tree Automata — Syntax

Σ ranked alphabet, I, Ω non-empty sets

- Tree representation over I, Σ, and Ω is $\mu = (\, \mu_k \mid k \in \mathbb{N}\,)$ such that

$$\mu_k : \Sigma_{(k)} \longrightarrow \Omega^{I \times I^k}$$

- $M = (I, \Sigma, \mathcal{D}, F, \mu)$ (bottom-up) DM-monoid weighted tree automaton (DM-wta), iff
  - I non-empty set of states,
  - Σ ranked alphabet of input symbols,
  - $\mathcal{D} = (D, \Omega, \sum)$ *complete DM-monoid*,
  - $F : I \longrightarrow \Omega_{(1)}$ final weight map, and
  - μ tree representation over I, Σ, and Ω such that $\mu_k : \Sigma_{(k)} \longrightarrow {\Omega_{(k)}}^{I \times I^k}$

DM-Monoid Weighted Tree Automata — Semantics

$\mathcal{D} = (D, \Omega, \sum)$ complete DM-monoid, $M = (I, \Sigma, \mathcal{D}, F, \mu)$ DM-wta.

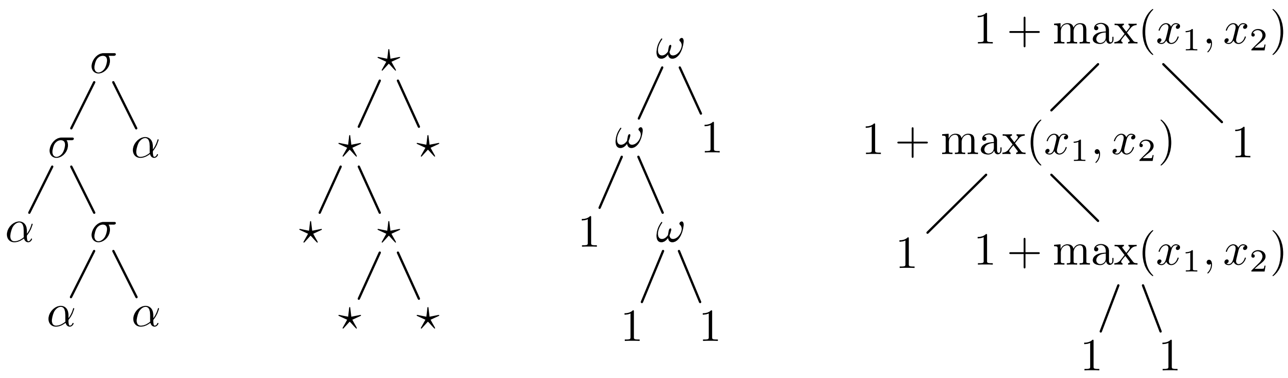
- Define $h_\mu : T_\Sigma \longrightarrow D^I$ by

$$h_\mu(\sigma(t_1, \ldots, t_k))_i = \sum_{i_1, \ldots, i_k \in I} \mu_k(\sigma)_{i,(i_1,\ldots,i_k)}\big(h_\mu(t_1)_{i_1}, \ldots, h_\mu(t_k)_{i_k}\big)$$

- $(\|M\|, t) = \sum_{i \in I} F_i(h_\mu(t)_i)$ is tree series recognized by M

Example DM-wta

- $\Sigma = \{\sigma, \alpha\}$ and $\Omega = \{\omega, \mathrm{id}, 1\}$ and $\omega(n_1, n_2) = 1 + \max(n_1, n_2)$,
- $\mathcal{N} = (\mathbb{N} \cup \{\infty\}, \Omega, \min)$ complete DM-monoid
- DM-wta $M_E = (\{\star\}, \Sigma, \mathcal{N}, F, \mu)$ with $F_\star = \mathrm{id}$, $\mu_0(\alpha)_\star = 1$, and $\mu_2(\sigma)_{\star,(\star,\star)} = \omega$



- $(\|M_E\|, t) = \mathrm{height}(t)$

Weighted Tree Automata & Tree Series Transducers

$M = (I, \Sigma, \mathcal{D}, F, \mu)$ DM-wta and $(A, \odot, \mathbf{0}, \mathbf{1}, \bigoplus)$ complete semiring

- M is weighted tree automaton (wta), iff $\mathcal{D} = (A, \Omega, \bigoplus)$ with $\Omega_{(k)} = \{\, \underline{a}_{(k)} \mid a \in A \,\}$ and

$$\underline{a}_{(k)}(d_1, \ldots, d_k) = a \odot d_1 \odot \cdots \odot d_k$$

- M is tree series transducer (tst), iff $\mathcal{D} = (A\langle\langle T_\Delta \rangle\rangle, \Omega, \bigoplus)$ with $\Omega_{(k)} = \{\, \underline{\psi}_{(k)} \mid \psi \in A\langle\langle T_\Delta(X_k) \rangle\rangle \,\}$ and

$$\underline{\psi}_{(k)}(\psi_1, \ldots, \psi_k) = \psi \longleftarrow (\psi_1, \ldots, \psi_k)$$

Constructing a Monoid (I)

$(D, \Omega, \sum)$ complete DM-monoid, $\Omega X = \{\, \overline{\omega}(x_1, \ldots, x_k) \mid k \in \mathbb{N}, \omega \in \Omega_{(k)} \,\}$

Theorem: There exists monoid $(B, \leftarrow, \varepsilon)$ such that $D \cup \Omega X \subseteq B$ and for all $d_1, \ldots, d_k \in D$

$$\omega(d_1, \ldots, d_k) = \overline{\omega}(x_1, \ldots, x_k) \leftarrow d_1 \leftarrow \cdots \leftarrow d_k.$$

Proof sketch: Let $\Omega' = \Omega \cup D$.

- Define $h : T_{\Omega'}(X) \longrightarrow T_{\Omega'}(X)$ for every $v \in D \cup X$ by

$$\begin{aligned} h(v) &= v \\ h(\overline{\omega}(t_1, \ldots, t_k)) &= \begin{cases} \omega(h(t_1), \ldots, h(t_k)) & \text{, if } h(t_1), \ldots, h(t_k) \in D \\ \overline{\omega}(h(t_1), \ldots, h(t_k)) & \text{, otherwise} \end{cases} \end{aligned}$$

- $\widehat{T_{\Omega'}}(X_n)$ set of X_n-contexts
- $h(t) \in \widehat{T_{\Omega'}}(X_n)$ iff $t \in \widehat{T_{\Omega'}}(X_n)$

Constructing a Monoid (II)

- Let $s(\!|t|\!) = s[t, x_{k+1}, x_{k+2}, \ldots, x_{k+n-1}]$ for $s \in \widehat{T_\Sigma}(X_n)$ and $t \in \widehat{T_\Sigma}(X_k)$ (non-identifying tree substitution).
- $B = D^* \cup \bigcup_{n \in \mathbb{N}_+} D^* \cdot \widehat{T_{\Omega'}}(X_n)$.
- Define $\leftarrow : B^2 \longrightarrow B$ for every $a \in D^*$, $b \in B$, $s \in \widehat{T_{\Omega'}}(X_n)$, $t \in D \cup \widehat{T_{\Omega'}}(X_n)$ by

$$
\begin{aligned}
a \leftarrow b &= a \cdot b \\
a \cdot s \leftarrow \varepsilon &= a \cdot s \\
a \cdot s \leftarrow t \cdot b &= a \cdot (h(s(\!|t|\!))) \leftarrow b.
\end{aligned}
$$

- $(B, \leftarrow, \varepsilon)$ is a monoid.
- $\omega(d_1, \ldots, d_k) = \overline{\omega}(x_1, \ldots, x_k) \leftarrow d_1 \leftarrow \cdots \leftarrow d_k$.

From a Monoid to a Semiring (I)

$\mathcal{A} = (A, \odot, \mathbf{0}, \mathbf{1}, \bigoplus)$ complete semiring, DM-monoid $(D, \Omega, \sum)$ complete semimodule of $\mathcal{A}$

- Lift mapping $\leftarrow : B^2 \longrightarrow B$ to a mapping $\leftarrow : A\langle\langle B\rangle\rangle^2 \longrightarrow A\langle\langle B\rangle\rangle$ by

$$\psi_1 \leftarrow \psi_2 = \bigoplus_{b_1, b_2 \in B} \big((\psi_1, b_1) \odot (\psi_2, b_2)\big)\,(b_1 \leftarrow b_2).$$

- Define sum of a series $\varphi \in A\langle\langle D\rangle\rangle$ (summed in D) by $\sum : A\langle\langle D\rangle\rangle \longrightarrow D$

$$\sum \varphi = \sum_{d \in D} (\varphi, d) \cdot d.$$

- **Theorem:**

  (i) $\sum(\bigoplus_{i\in I} \varphi_i) = \sum_{i \in I} \sum \varphi_i$ for every family $(\, \varphi_i \mid i \in I \,)$ of series and

  (ii) $\omega(\sum \varphi_1, \ldots, \sum \varphi_k) = \sum\big(\overline{\omega}(x_1, \ldots, x_k) \leftarrow \varphi_1 \leftarrow \cdots \leftarrow \varphi_k\big).$

From a Monoid to a Semiring (II)

$(D, \Omega, \sum)$ continuous DM-monoid, $M_1 = (I, \Sigma, \mathcal{D}, F_1, \mu_1)$ DM-wta.

- Theorem: There exists complete semiring $(C, \leftarrow, \widetilde{\mathbf{0}}, \varepsilon, \bigoplus)$ such that $D \cup \Omega X \subseteq C$ and

  (i) $\omega(d_1, \ldots, d_k) = \overline{\omega}(x_1, \ldots, x_k) \leftarrow d_1 \leftarrow \cdots \leftarrow d_k$,

  (ii) $\sum(\bigoplus_{i \in I} d_i) = \sum_{i \in I} d_i$.

  Proof sketch: Let $(A, \odot, \mathbf{0}, \mathbf{1}, \bigoplus)$ complete semiring such that $\mathcal{D}$ is a complete semimodule thereof. There exists monoid $(B, \leftarrow, \varepsilon)$ such that (i) holds. Let $C = A\langle\langle B \rangle\rangle$ and $\leftarrow : C^2 \longrightarrow C$ be the extension of $\leftarrow$ on B.

- Theorem: There exists a wta $M = (I, \Sigma, \mathcal{B}, F, \mu)$ such that $\|M_1\| = \sum \|M\|$.

Establishing a Relationship

- Theorem: For every tst M_1, there exists a wta M such that $\sum \|M\| = \|M_1\|$.
- Theorem: For every deterministic tst M_2, there exists a deterministic wta M such that $\|M\| = \|M_2\|$.
- Theorem: For every tree transducer M_3, there exists a wta M such that $\|M\| = \|M_3\|$.
- Theorem: For every tst M_4 over an $\aleph_0$-idempotent semiring, there exists a wta M such that $\|M\| = \|M_4\|$.

Pumping Lemma for DM-wta

$\mathcal{D} = (D, \Omega, \sum)$ complete DM-monoid, $L \in \mathcal{L}^d_\Sigma(\mathcal{D})$, and $\Omega' = \Omega \cup D$.

Theorem: There exists $m \in \mathbb{N}$ such that for every $t \in \mathrm{supp}(L)$ with $\mathrm{height}(t) \geq m + 1$ there exist $C, C' \in \widehat{T_\Sigma}(X_1)$, $s \in T_\Sigma$, and $a, a' \in \widehat{T_{\Omega'}}(X_1)$, and $d \in D$ such that

- $t = C[C'[s]]$,
- $\mathrm{height}(C[s]) \leq m + 1$ and $C \neq x_1$, and
- $(L, C'[C^n[s]]) = a' \leftarrow a^n \leftarrow d$ for every $n \in \mathbb{N}$.

Conclusions

- the study of arbitrary weighted tree automata provides results for tree series transducers
- e.g., a pumping lemma for tree series transducers can be derived from a pumping lemma for weighted tree automata
- unfortunately, few results for weighted tree automata over non-commutative semirings exist

Thank You for Your Attention.