Determinization of Weighted Tree Automata using Factorizations

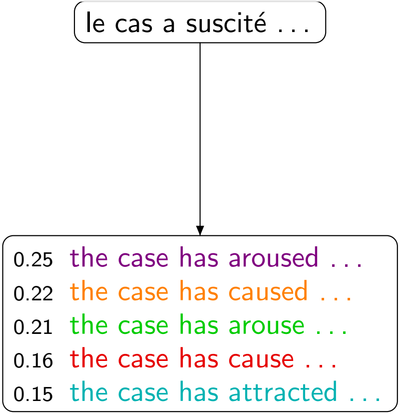
Matthias Büchse   Jonathan May   Heiko Vogler

2009-09-01

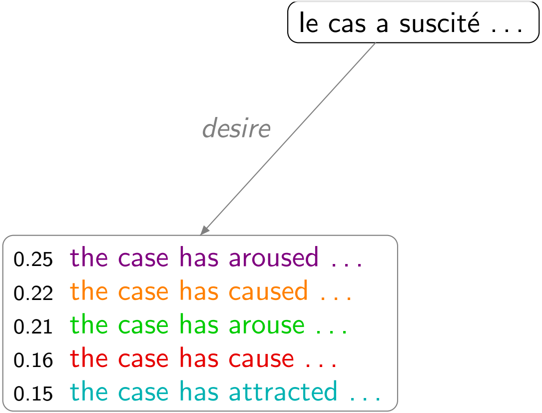
Machine Translation (MT)

le cas a suscité . . .

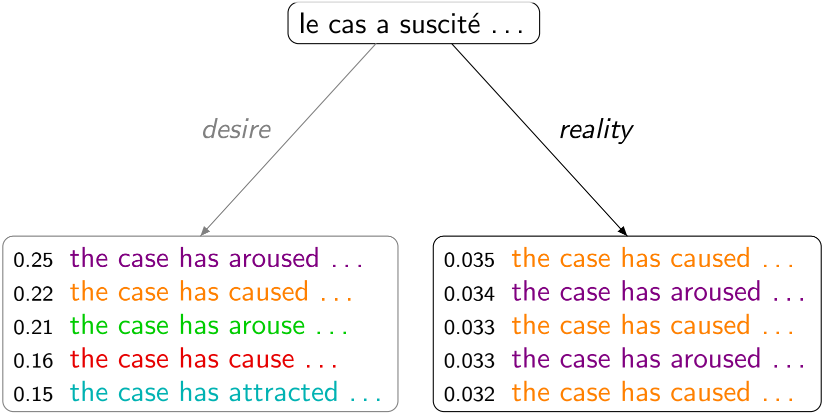
Machine Translation (MT)



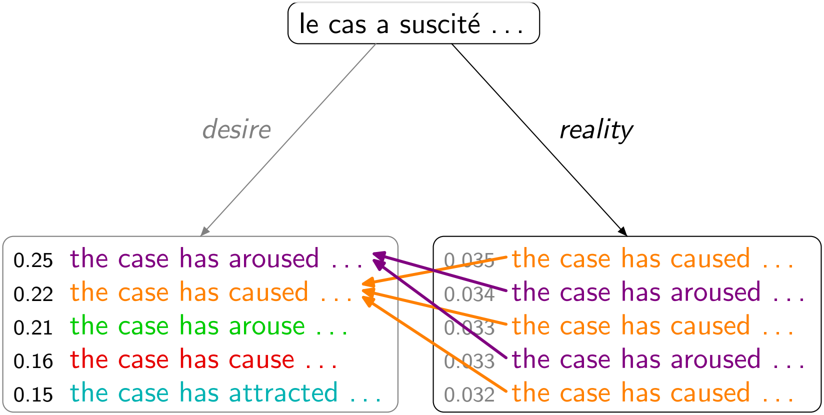
Machine Translation (MT)



Machine Translation (MT)



Machine Translation (MT)



Bridging the gap

Actual output: list of best derivations

0.035: the case has caused . . .
0.034: the case has aroused . . .
0.033: the case has caused . . .
0.033: the case has aroused . . .
0.032: the case has caused . . .

- merge derivations in the output list

Bridging the gap

Modified actual output

0.1: the case has caused . . .
0.034: the case has aroused . . .
0.033: the case has aroused . . .

Bridging the gap

Modified actual output

0.1: the case has caused . . .
0.067: the case has aroused . . .

Bridging the gap

Modified actual output

0.1: the case has caused ...
0.067: the case has aroused ...
0.032: the case has caused ...
0.032: the case has aroused ...
0.032: the case has aroused ...
0.031: the case has arouse ...
0.03: the case has aroused ...

Bridging the gap

Modified actual output

0.132: the case has caused . . .
0.067: the case has aroused . . .
0.032: the case has aroused . . .
0.032: the case has aroused . . .
0.031: the case has arouse . . .
0.03: the case has aroused . . .

Bridging the gap

Modified actual output

0.161: the case has aroused ...

0.132: the case has caused ...

0.031: the case has arouse ...

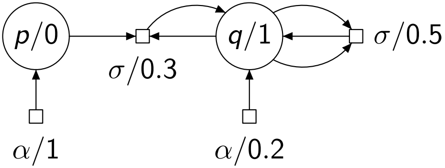
Bridging the gap

Modified actual output

0.161: the case has aroused . . .
0.132: the case has caused . . .
0.031: the case has arouse . . .

- still more derivations needed!
- actual weights unknown
- reengineer the system

MT system ≈ weighted tree automaton (wta)

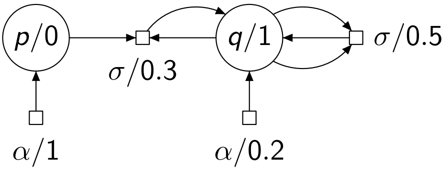


Definition of wta

$M = (Q, \mu, \nu)$ wta (over Σ and A)

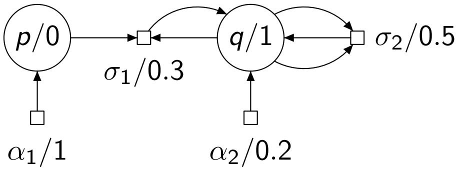
- Q is a nonempty finite set (of states),
- μ is a family $(\mu_k \mid k \in \mathbb{N})$ (of transition mappings) with $\mu_k : \Sigma^{(k)} \to A^{Q^k \times Q}$ for every $k \in \mathbb{N}$, and
- $\nu \in A^Q$ (final weights).

Unfoldings in wta I



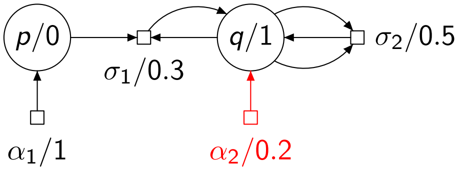
10 best unfoldings:

Unfoldings in wta I



10 best unfoldings:

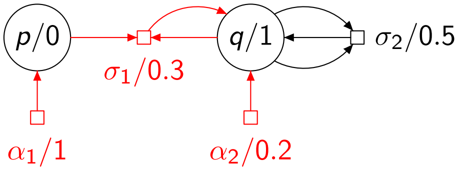
Unfoldings in wta I



10 best unfoldings:

α_2 0.2

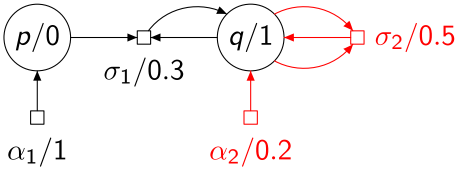
Unfoldings in wta I



10 best unfoldings:

α_2	0.2
$\sigma_1(\alpha_1, \alpha_2)$	0.06

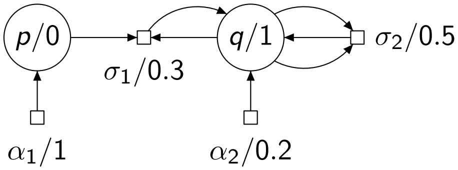
Unfoldings in wta I



10 best unfoldings:

α_2	0.2
$\sigma_1(\alpha_1, \alpha_2)$	0.06
$\sigma_2(\alpha_2, \alpha_2)$	0.02

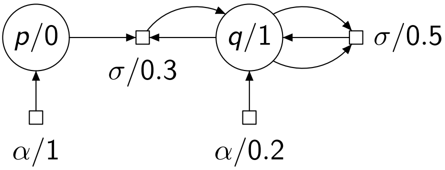
Unfoldings in wta I



10 best unfoldings:

α_2	0.2
$\sigma_1(\alpha_1, \alpha_2)$	0.06
$\sigma_2(\alpha_2, \alpha_2)$	0.02
$\sigma_1(\alpha_1, \sigma_1(\alpha_1, \alpha_2))$	0.018
$\sigma_1(\alpha_1, \sigma_2(\alpha_2, \alpha_2))$	0.006
$\sigma_2(\alpha_2, \sigma_1(\alpha_1, \alpha_2))$	0.006
$\sigma_2(\sigma_1(\alpha_1, \alpha_2), \alpha_2)$	0.006
$\sigma_1(\alpha_1, \sigma_1(\alpha_1, \sigma_1(\alpha_1, \alpha_2)))$	0.0054
$\sigma_2(\alpha_2, \sigma_2(\alpha_2, \alpha_2))$	0.002
$\sigma_2(\sigma_2(\alpha_2, \alpha_2), \alpha_2)$	0.002

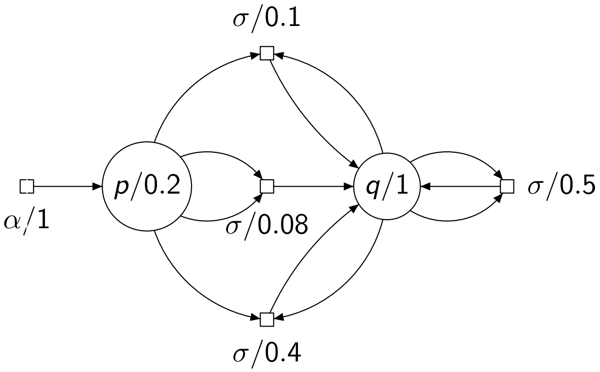
Unfoldings in wta I



10 best unfoldings:

α	0.2
$\sigma(\alpha, \alpha)$	0.06
$\sigma(\alpha, \alpha)$	0.02
$\sigma(\alpha, \sigma(\alpha, \alpha))$	0.018
$\sigma(\alpha, \sigma(\alpha, \alpha))$	0.006
$\sigma(\alpha, \sigma(\alpha, \alpha))$	0.006
$\sigma(\sigma(\alpha, \alpha), \alpha)$	0.006
$\sigma(\alpha, \sigma(\alpha, \sigma(\alpha, \alpha)))$	0.0054
$\sigma(\alpha, \sigma(\alpha, \alpha))$	0.002
$\sigma(\sigma(\alpha, \alpha), \alpha)$	0.002

Unfoldings in wta ii

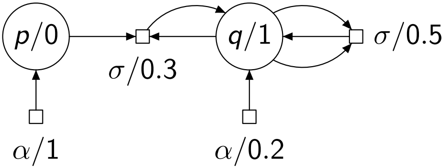


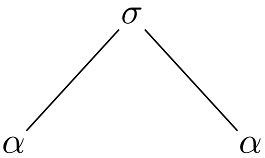
- list of best unfoldings matches list of best trees

Deterministic wta

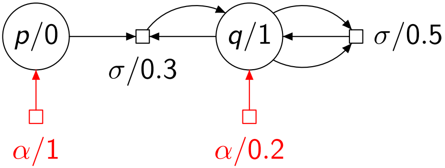
- A wta is called deterministic if every tree can be unfolded from at most one state.
- If a wta is deterministic, it allows at most one unfolding for every tree.
- Then: list of best trees and list of best derivations coincide.
- Determinization: Try to construct equivalent deterministic wta.

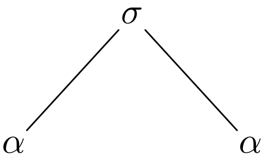
Semantics: example



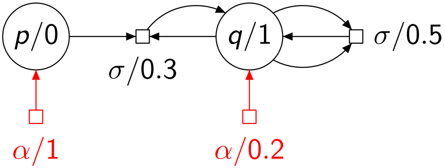


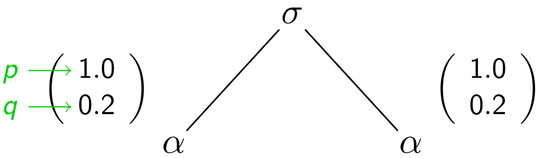
Semantics: example



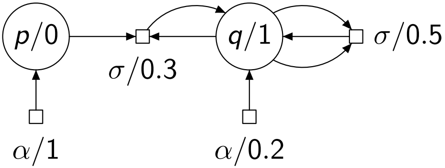


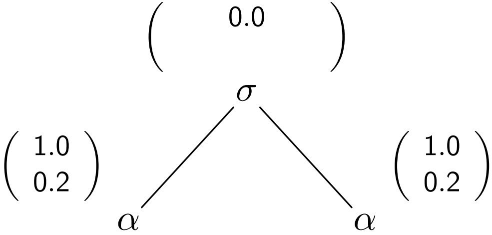
Semantics: example



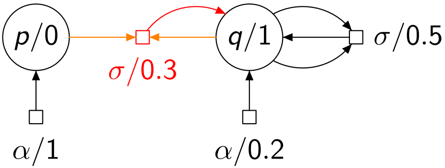


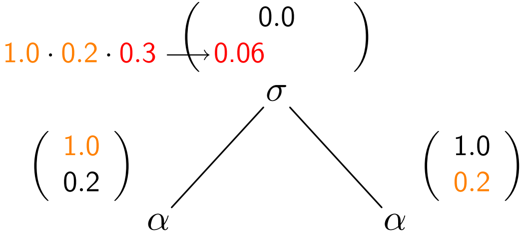
Semantics: example



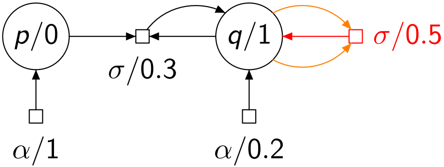


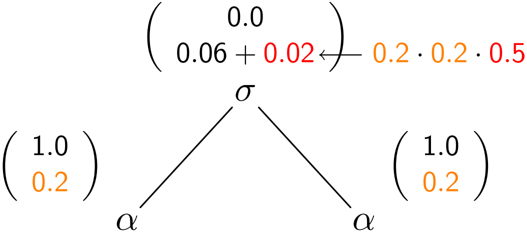
Semantics: example



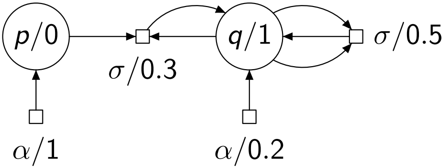


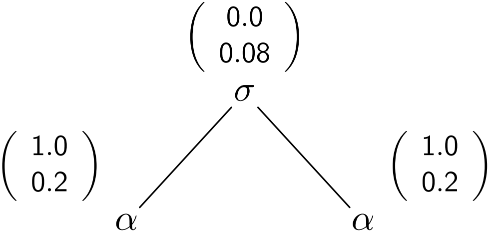
Semantics: example



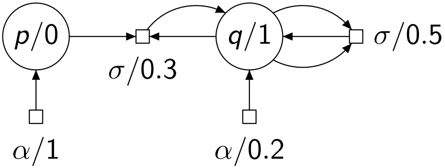


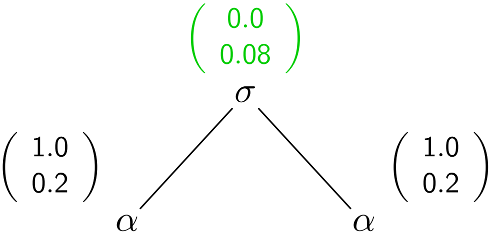
Semantics: example





Semantics: example

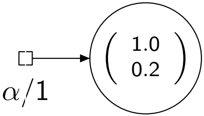




every tree has a uniquely determined weight vector!

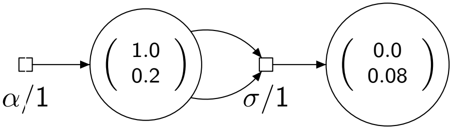
Determinization

according to Borchardt (2004)



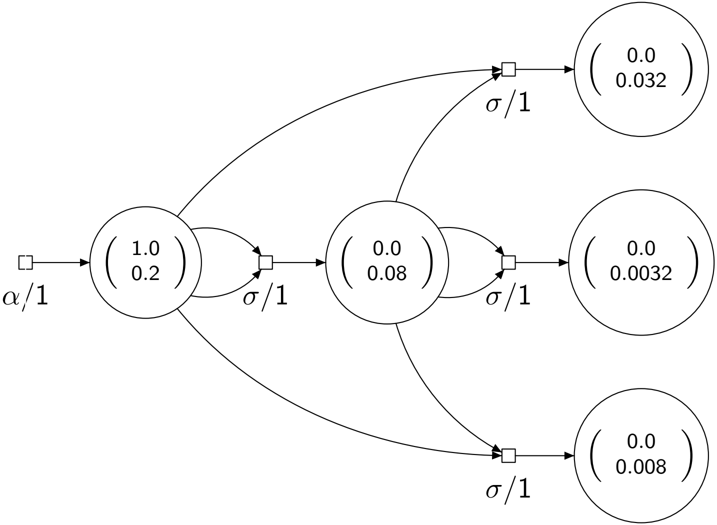
Determinization

according to Borchardt (2004)



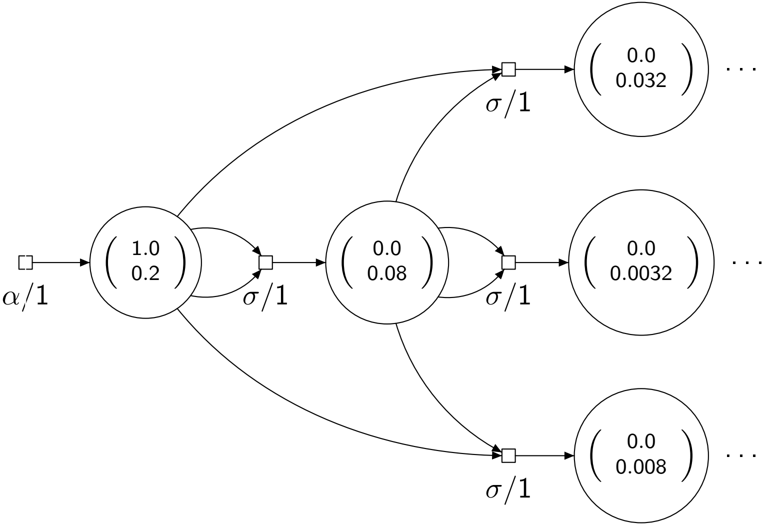
Determinization

according to Borchardt (2004)

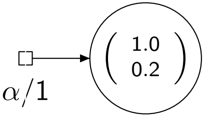


Determinization

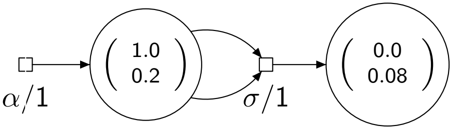
according to Borchardt (2004)



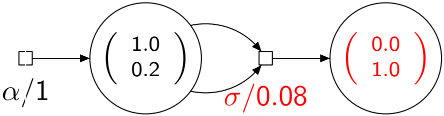
Introducing factorizations



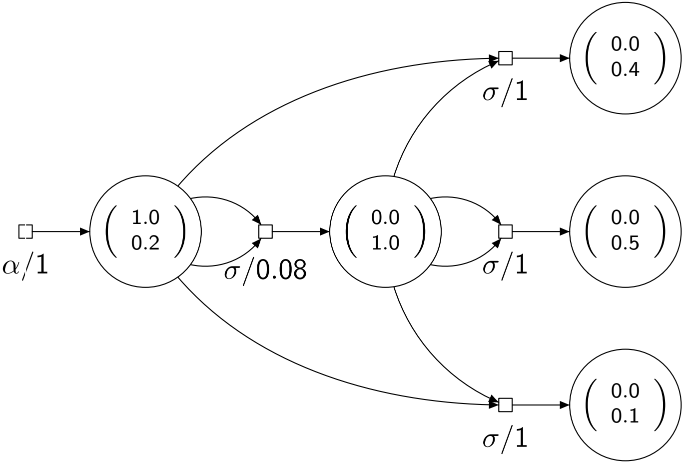
Introducing factorizations



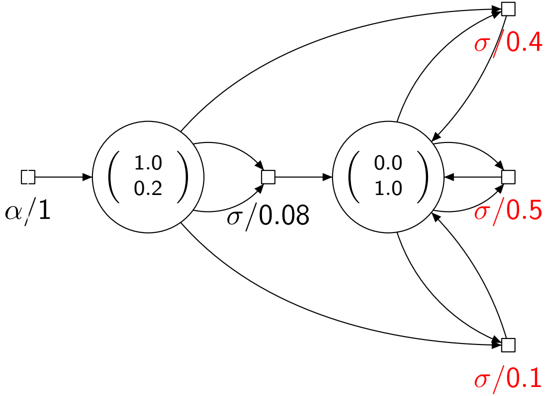
Introducing factorizations



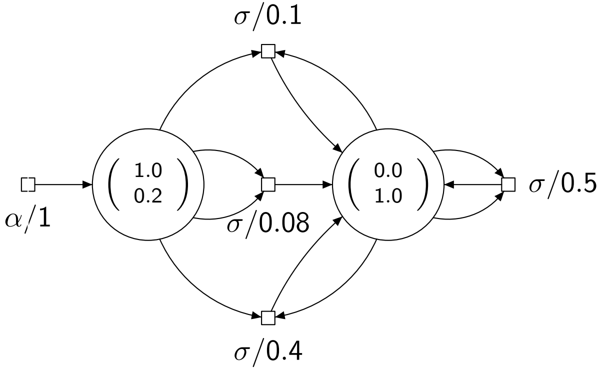
Introducing factorizations



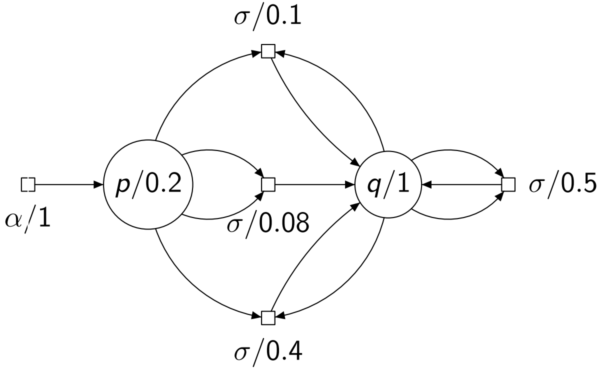
Introducing factorizations



Introducing factorizations



Introducing factorizations



History of factorizations

History of factorizations

- Mohri 1997

History of factorizations

- Mohri 1997
- Kirsten and Mäurer 2005

History of factorizations

- Mohri 1997
- Kirsten and Mäurer 2005
- May and Knight 2006

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and
- 0 is absorbing wrt $\cdot$.

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and
- 0 is absorbing wrt $\cdot$.

Example

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and
- 0 is absorbing wrt $\cdot$.

Example

- natural numbers: $(\mathbb{N}, +, \cdot, 0, 1)$,

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and
- 0 is absorbing wrt $\cdot$.

Example

- natural numbers: $(\mathbb{N}, +, \cdot, 0, 1)$,
- Tropical semiring: $(\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$,

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and
- 0 is absorbing wrt $\cdot$.

Example

- natural numbers: $(\mathbb{N}, +, \cdot, 0, 1)$,
- Tropical semiring: $(\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$,
- Viterbi semiring: $([0, 1], \max, \cdot, 0, 1)$,

Weight domains: commutative semirings

Definition

A commutative semiring is a quintuple $(A, +, \cdot, 0, 1)$ such that

- A is a set,
- $+$ and $\cdot$ are associative, commutative binary operations on A,
- 0 and 1 are neutral elements of $+$ and $\cdot$, resp.,
- $\cdot$ distributes over $+$, and
- 0 is absorbing wrt $\cdot$.

Example

- natural numbers: $(\mathbb{N}, +, \cdot, 0, 1)$,
- Tropical semiring: $(\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$,
- Viterbi semiring: $([0, 1], \max, \cdot, 0, 1)$,
- non-negative reals: $(\mathbb{R}^{\geq 0}, +, \cdot, 0, 1)$.

Factorizations

(Kirsten and Mäurer 2005)

Definition

A pair (f, g) is a factorization if

1. $f : A^Q \setminus \{0^Q\} \to A^Q$,
2. $g : A^Q \setminus \{0^Q\} \to A$, and
3. $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{0^Q\}$.

Factorizations

(Kirsten and Mäurer 2005)

Definition

A pair (f, g) is a factorization if

1. $f : A^Q \setminus \{0^Q\} \to A^Q$,
2. $g : A^Q \setminus \{0^Q\} \to A$, and
3. $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{0^Q\}$.

Example

Tropical semiring: $A = (\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$.

- $g(u) = \min\{u_q \mid q \in Q\}$,
- $f(u)_q = u_q - g(u)$.

Factorizations

(Kirsten and Mäurer 2005)

Definition

A pair (f, g) is a factorization if

1. $f : A^Q \setminus \{0^Q\} \to A^Q$,
2. $g : A^Q \setminus \{0^Q\} \to A$, and
3. $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{0^Q\}$.

Example

Tropical semiring: $A = (\mathbb{R}^{\geq 0}_{\infty}, \min, +, \infty, 0)$.

- $g(u) = \min\{u_q \mid q \in Q\}$,
- $f(u)_q = u_q - g(u)$.

$$g(\begin{pmatrix} 2 \\ 3 \end{pmatrix}) = 2, \quad f(\begin{pmatrix} 2 \\ 3 \end{pmatrix}) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Factorizations

(Kirsten and Mäurer 2005)

Definition

A pair (f, g) is a factorization if

1. $f : A^Q \setminus \{0^Q\} \to A^Q$,
2. $g : A^Q \setminus \{0^Q\} \to A$, and
3. $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{0^Q\}$.

Example

non-negative reals: $A = (\mathbb{R}^{\geq 0}, +, \cdot, 0, 1)$.

- $g(u) = \sum_{q \in Q} u_q$,
- $f(u)_q = u_q \cdot g(u)^{-1}$.

Factorizations

(Kirsten and Mäurer 2005)

Definition

A pair (f, g) is a factorization if

1. $f : A^Q \setminus \{0^Q\} \to A^Q$,
2. $g : A^Q \setminus \{0^Q\} \to A$, and
3. $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{0^Q\}$.

Example

non-negative reals: $A = (\mathbb{R}^{\geq 0}, +, \cdot, 0, 1)$.

- $g(u) = \sum_{q \in Q} u_q$,
- $f(u)_q = u_q \cdot g(u)^{-1}$.

$$g(\begin{pmatrix} 2 \\ 3 \end{pmatrix}) = 5, \quad f(\begin{pmatrix} 2 \\ 3 \end{pmatrix}) = \begin{pmatrix} 0.4 \\ 0.6 \end{pmatrix}$$

Factorizations

(Kirsten and Mäurer 2005)

Definition

A pair (f, g) is a factorization if

1. $f : A^Q \setminus \{0^Q\} \to A^Q$,
2. $g : A^Q \setminus \{0^Q\} \to A$, and
3. $u = g(u) \cdot f(u)$ for every $u \in A^Q \setminus \{0^Q\}$.

Definition

A factorization (f, g) is maximal if

$$f(u) = f(a \cdot u)$$

for every $a \in A$ and $u \in A^Q$ with $a \cdot u \neq 0^Q$.

Construction of deterministic wta

Construction of deterministic wta

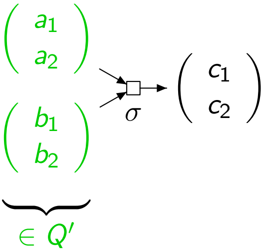
1. start with empty automaton, $Q' = \emptyset$

Construction of deterministic wta

1. start with empty automaton, $Q' = \emptyset$
2. close off under:

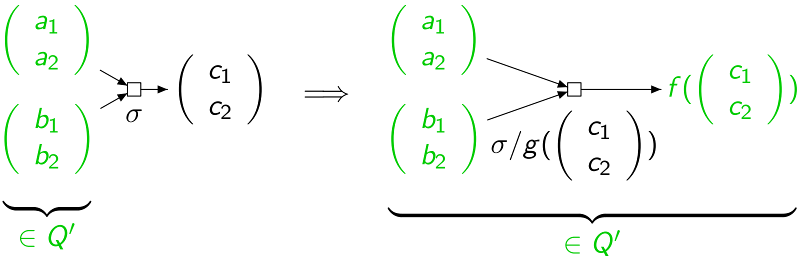
Construction of deterministic wta

1. start with empty automaton, $Q' = \emptyset$
2. close off under:



Construction of deterministic wta

1. start with empty automaton, $Q' = \emptyset$
2. close off under:



Formal construction

$\det_{(f,g)}(M) = (Q', \mu', \nu')$ where

- Q' is the smallest set $P \subseteq A^Q$ such that for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in P$: if $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$, then $f(\mu_M(\sigma)(u_1, \ldots, u_k)) \in P$.

Formal construction

$\det_{(f,g)}(M) = (Q', \mu', \nu')$ where

- Q' is the smallest set $P \subseteq A^Q$ such that for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in P$: if $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$, then $f(\mu_M(\sigma)(u_1, \ldots, u_k)) \in P$.
- $\mu'_k(\sigma)_{u_1 \ldots u_k, u} =$

$$\begin{cases} g(\mu_M(\sigma)(u_1, \ldots, u_k)) & \text{if } \mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0} \text{ and} \\ & \quad u = f(\mu_M(\sigma)(u_1, \ldots, u_k)) , \\ 0 & \text{otherwise ,} \end{cases}$$

for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in Q'$, and $u \in Q'$,

Formal construction

$\det_{(f,g)}(M) = (Q', \mu', \nu')$ where

- Q' is the smallest set $P \subseteq A^Q$ such that for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in P$: if $\mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0}$, then $f(\mu_M(\sigma)(u_1, \ldots, u_k)) \in P$.
- $\mu'_k(\sigma)_{u_1 \ldots u_k, u} =$
$$\begin{cases} g(\mu_M(\sigma)(u_1, \ldots, u_k)) & \text{if } \mu_M(\sigma)(u_1, \ldots, u_k) \neq \tilde{0} \text{ and} \\ & \quad u = f(\mu_M(\sigma)(u_1, \ldots, u_k)) \,, \\ 0 & \text{otherwise} \,, \end{cases}$$
for every $k \in \mathbb{N}$, $\sigma \in \Sigma^{(k)}$, $u_1, \ldots, u_k \in Q'$, and $u \in Q'$,
- $\nu'_u = \sum_{q \in Q} u_q \cdot \nu_q$ for every $u \in Q'$.

Results and problems

Results

- determinization construction for wta over commutative semirings
- resulting automaton not necessarily finite
- correctness
- sufficient conditions for finiteness:
  maximal factorization and
  - nonrecursive wta (acyclic hypergraph) or
  - extremal semiring, twins property (Kirsten and Mäurer 2005)

Results and problems

Results

- determinization construction for wta over commutative semirings
- resulting automaton not necessarily finite
- correctness
- sufficient conditions for finiteness:
  maximal factorization and
  - nonrecursive wta (acyclic hypergraph) or
  - extremal semiring, twins property (Kirsten and Mäurer 2005)

Extremal semirings

$$a + b \in \{a, b\}$$

Results and problems

Results

- determinization construction for wta over commutative semirings
- resulting automaton not necessarily finite
- correctness
- sufficient conditions for finiteness:
  maximal factorization and
  - nonrecursive wta (acyclic hypergraph) or
  - extremal semiring, twins property (Kirsten and Mäurer 2005)

Extremal semirings

$$a + b \in \{a, b\}$$

only the best derivation counts

Results and problems

Results

- determinization construction for wta over commutative semirings
- resulting automaton not necessarily finite
- correctness
- sufficient conditions for finiteness:
  maximal factorization and
  - nonrecursive wta (acyclic hypergraph) or
  - extremal semiring, twins property (Kirsten and Mäurer 2005)

Twins property

undecidable in general

Twins property

A wta $M = (Q, \mu, \nu)$ is said to have the twins property if for every $p, q \in Q$, $\xi \in T_\Sigma$, and $\zeta \in C_\Sigma$, we have that if $\mathrm{wt}_M(R_M^p(\zeta \cdot \xi)) \neq 0$ and $\mathrm{wt}_M(R_M^q(\zeta \cdot \xi)) \neq 0$, then $\mathrm{wt}_M(R_M^{p,p}(\zeta)) = \mathrm{wt}_M(R_M^{q,q}(\zeta))$.

References

B. Borchardt.
A pumping lemma and decidability problems for recognizable tree series.
Acta Cybernet., 16(4):509–544, 2004.

D. Kirsten and I. Mäurer.
On the determinization of weighted automata.
J. Autom. Lang. Comb., 10:287–312, 2005.

J. May and K. Knight.
A better N-best list: practical determinization of weighted finite tree automata.
In *Proc. NAACL-HLT*, 2006.

M. Mohri.
Finite-state transducers in language and speech processing.
Computational Linguistics, 23(2):269–311, 1997.