An Automata Theoretic Approach to Rational Tree Relations

Frank Radmacher

RWTH AACHEN

SOFSEM 2008

Motivation: Rational Relations over Words

There exist several equivalent descriptions of rational word relations, e. g.

- rational expressions:
  - for instance $R = \{(v, uv) \mid u, v \in \Sigma^*\}$ is definable by $\left(\binom{\varepsilon}{a} + \binom{\varepsilon}{b}\right)^* \cdot \left(\binom{a}{a} + \binom{b}{b}\right)^*$
- asynchronous automata (also called multitape automata):
  - (these have transitions of the form $p \xrightarrow{a_1/\ldots/a_n} q$ in $Q \times (\Sigma_1 \cup \{\varepsilon\}) \times \ldots \times (\Sigma_n \cup \{\varepsilon\}) \times Q$)

Rational Tree Relations

- Goal: Generalization to trees with the alternative descriptions by rational expressions and asynchronous automata.
- Starting point: Definition via rational expressions over trees (J.-C. Raoult (1997)).
- An automata theoretic approach is still missing.
  We define asynchronous tree automata recognizing exactly Raoult's relations.

Rational Tree Expressions

Jean-Claude Raoult.
Rational tree relations.
Bull. Belg. Math. Soc., 4(1): 149–176, 1997.

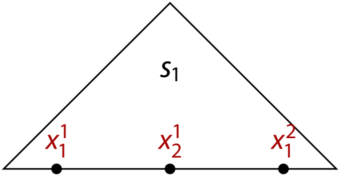
Definition

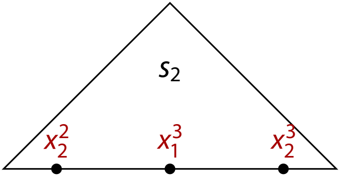
The classes Rat_n of rational tree relations are defined inductively:

- Each finite n-ary tree relation is in Rat_n.
- $R \in \mathrm{Rat}_n \wedge S \in \mathrm{Rat}_n \implies R \cup S \in \mathrm{Rat}_n$.
- $R \in \mathrm{Rat}_n \wedge |X| = m \wedge S \in \mathrm{Rat}_m \implies R \cdot_X S \in \mathrm{Rat}_n$.
- $R \in \mathrm{Rat}_n \wedge |X| = n \implies R^{*_X} \in \mathrm{Rat}_n$.

Concatenation via a Multivariable X

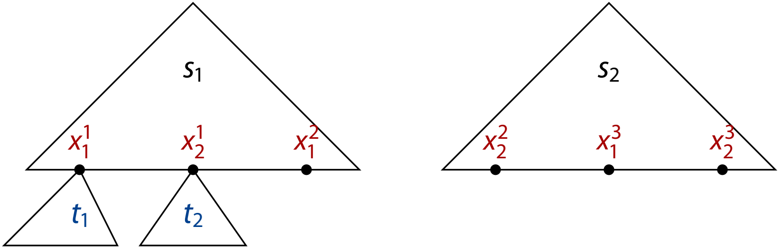
$\{(s_1, s_2)\} \cdot_X \{(t_1, t_2), (t'_1, t'_2)\}$ with respect to a multivariable $X = x_1 x_2$:





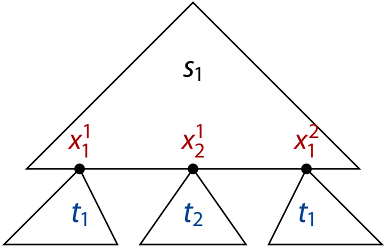
Concatenation via a Multivariable X

$\{(s_1, s_2)\} \cdot_X \{(t_1, t_2), (t_1', t_2')\}$ with respect to a multivariable $X = x_1 x_2$:



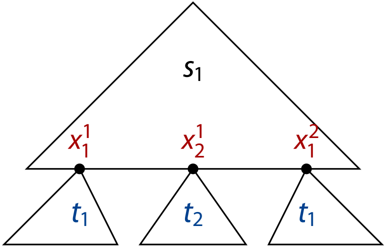
Concatenation via a Multivariable X

$\{(s_1, s_2)\} \cdot_X \{(t_1, t_2), (t'_1, t'_2)\}$ with respect to a multivariable $X = x_1 x_2$:



Concatenation via a Multivariable X

$\{(s_1, s_2)\} \cdot_X \{(t_1, t_2), (t'_1, t'_2)\}$ with respect to a multivariable $X = x_1 x_2$:

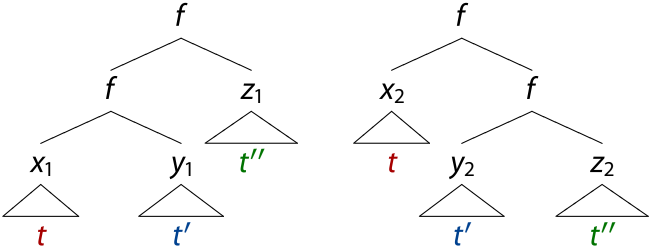


($2^3 = 8$ possibilities)

Example 1

Rational expression

$(ffx_1y_1z_1, fx_2fy_2z_2) \cdot_{x_1x_2} (t, t) \cdot_{y_1y_2} (t', t') \cdot_{z_1z_2} (t'', t'')$

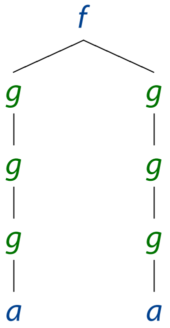


Right rotation is definable by a rational expression.

Example 2

Rational expression

$(f x_1 x_2) \cdot_{x_1 x_2} (g x_1, g x_2)^{*_{x_1 x_2}} \cdot_{x_1 x_2} (a, a)$

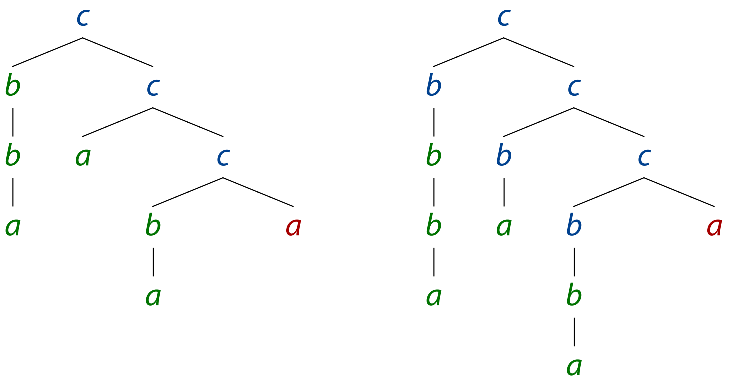


Consequence: Unary rational tree relations do not coincide with regular tree languages.

Example 3

Rational expression

$(cx_1y_1, cbx_2y_2)^{*_{y_1y_2}} \cdot_{y_1y_2} (a, a) \cdot_{x_1x_2} (bx_1, bx_2)^{*_{x_1x_2}} \cdot_{x_1x_2} (a, a)$



An unbounded number of multivariable instances can be generated.

Asynchronous Tree Automata

What do we need for an automata theoretic approach?

Asynchronous Tree Automata

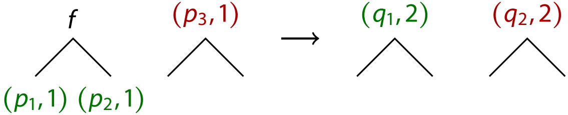
What do we need for an automata theoretic approach?

- Mechanism analogous to multivariables.
  - Macro states: e. g. $\mathfrak{p} = (p_1, p_2, p_3)$, $\mathfrak{q} = (q_1, q_2)$
  - Finite set of macro states: e. g. $\mathfrak{Q} = \{\mathfrak{p}, \mathfrak{q}\}$

Asynchronous Tree Automata

What do we need for an automata theoretic approach?

- Mechanism analogous to multivariables.
  - Macro states: e. g. $\mathfrak{p} = (p_1, p_2, p_3)$, $\mathfrak{q} = (q_1, q_2)$
  - Finite set of macro states: e. g. $\mathfrak{Q} = \{\mathfrak{p}, \mathfrak{q}\}$
- Instances of macro states have to be distinguishable.
  - Combining states with formal indices in the transitions:
    $((p_1, i), (p_2, i), f, (q_1, j))$, $((p_3, i), \varepsilon, (q_2, j))$
  - Instantiation in the bottom-up run with natural numbers:



  - Only complete instances of macro states are left and reached.

Decidability Results

- The following problems are decidable:
  - membership problem
  - emptiness problem (via reachability of final macro states)

Decidability Results

- The following problems are decidable:
  - membership problem
  - emptiness problem (via reachability of final macro states)

- Undecidability Results are inherited from rational word relations.
  - $R \cap S = \emptyset$
  - $R \subseteq S$
  - $R = S$

  It is undecidable whether a rational tree language is regular.

Drawbacks

- Restricted to unary sets, rational tree relations do not coincide with the class of regular tree languages.

- Non-closure under composition (for binary relations).

- When a binary relation $R \subseteq A \times B$ is considered as transduction $\tau_R : A \to \mathcal{P}(B)$, rational tree relations do not preserve regular tree languages.

  (e. g. the image over $T_\Sigma \times \{f(g^n a, g^n a) \mid n \in \mathbb{N}\}$ is not regular.)

Restrictions of Rational Tree Relations

Restrictions of rational tree relations to overcome the drawbacks:

- Transduction Grammars (J.-C. Raoult)
  - + Still good expressiveness
  - – Not a proper generalization of n-ary rational relations over words
  - – Restriction misses a natural automata theoretic description in our framework

Restrictions of Rational Tree Relations

Restrictions of rational tree relations to overcome the drawbacks:

- Transduction Grammars (J.-C. Raoult)
  - \+ Still good expressiveness
  - – Not a proper generalization of *n*-ary rational relations over words
  - – Restriction misses a natural automata theoretic description in our framework
- Separate-Rational Tree Relations (new proposal)
  - \+ Restriction is also natural for asynchronous tree automata
  - \+ Generalize *n*-ary rational word relations
  - – Lack definability of some important relations and classes, e. g. left/right rotations, linear tree transducers

Both generalize automatic tree relations.

Separate-Rational Tree Relations

Definition

The classes SepRat_n of separate-rational tree relations are defined inductively:

- Each finite n-ary tree relation is in SepRat_n
- $R \in \mathrm{SepRat}_n \ \wedge \ |X| = m \ \wedge \ S \in \mathrm{SepRat}_m$
  $\Longrightarrow \ R \cup S \in \mathrm{SepRat}_n, \ R \cdot_X S \in \mathrm{SepRat}_n, \ R^{*_X} \in \mathrm{SepRat}_n$

Separate-Rational Tree Relations

Definition

The classes SepRat_n of separate-rational tree relations are defined inductively:

- Each finite n-ary tree relation is in SepRat_n
  - which only consists of trees with a height of 1 or 2 nodes
  - and has at most one variable of each multivariable in each component.
- $R \in \text{SepRat}_n \ \wedge\ |X| = m \ \wedge\ S \in \text{SepRat}_m$
  $\Longrightarrow\ R \cup S \in \text{SepRat}_n,\ R \cdot_X S \in \text{SepRat}_n,\ R^{*_X} \in \text{SepRat}_n$
  - where each component of a a tuple in R contains at most one variable of X.

Separate-Asynchronous Tree Automata

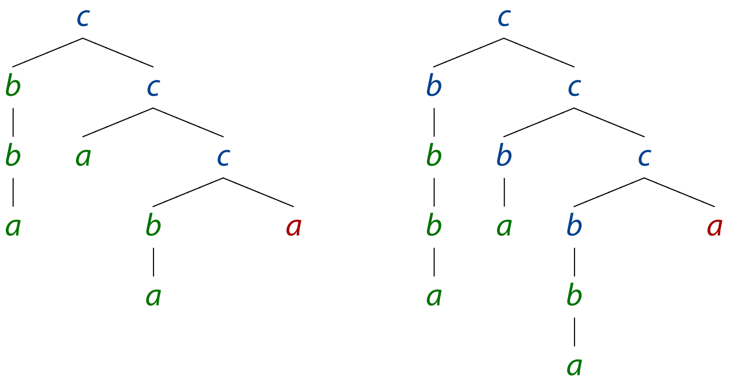
The restriction is also natural for asynchronous tree automata:

- Partition of the state set: Every state q_i of a macro state $(q_1, \ldots, q_m)$ can occur in a run only in one projection.
- Consequence: Synchronization within one projection of the relation is prevented.

Example 3 revisited

Rational expression

$(cx_1y_1, cx_2y_2)^{*_{y_1y_2}} \cdot_{x_1x_2} (x_1, bx_2) \cdot_{y_1y_2} (a, a) \cdot_{x_1x_2} (bx_1, bx_2)^{*_{x_1x_2}} \cdot_{x_1x_2} (a, a)$

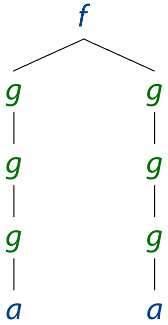


Relation is also separate-rational.

Example 2 revisited

Rational expression

$(f x_1 x_2) \cdot_{x_1 x_2} (g x_1, g x_2)^{*_{x_1 x_2}} \cdot_{x_1 x_2} (a, a)$

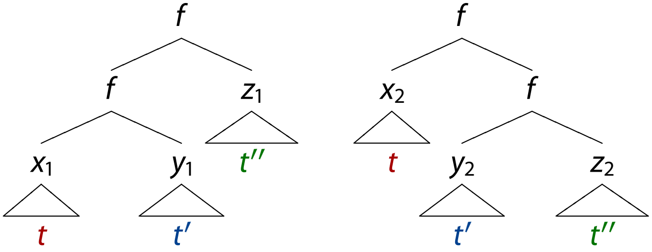


Definable by an rational expression, but not by a separate-rational.

Example 1 revisited

Rational expression

$(ffx_1y_1z_1, fx_2fy_2z_2) \cdot_{x_1x_2} (t, t) \cdot_{y_1y_2} (t', t') \cdot_{z_1z_2} (t'', t'')$



Right rotation is not separate-rational.

Conclusion

- Summary:
  - Rational tree relations are a non-trivial generalization of rational relations over words.
  - Asynchronous tree automata enable further investigation of the theory.
  - Restrictions are required to preserve the good properties of rational word relations.

Conclusion

- Summary:
  - Rational tree relations are a non-trivial generalization of rational relations over words.
  - Asynchronous tree automata enable further investigation of the theory.
  - Restrictions are required to preserve the good properties of rational word relations.

- Outlook:

  The automata theoretic approach enables the definition of
  - rational relations over unranked trees
  - deterministic rational tree relations

Conclusion

- Summary:
  - Rational tree relations are a non-trivial generalization of rational relations over words.
  - Asynchronous tree automata enable further investigation of the theory.
  - Restrictions are required to preserve the good properties of rational word relations.

- Outlook:

  The automata theoretic approach enables the definition of
  - rational relations over unranked trees
  - deterministic rational tree relations
    - a top-down approach is straightforward
    - the bottom-up approach seems to be challenging