Resynchronizing Classes of Word Relations

María Emilia Descotte
LaBRI [1]

DELTA Meeting in Paris
March 26th 2018

[1] Joint work with D. Figueira and G. Puppis

Synchronized pairs of words (over a fixed alphabet $\mathbb{A}$)

Synchronizing pairs of words

A **synchronization** of (w_1, w_2) is a word over $\mathbf{2} \times \mathbb{A}$ so that the projection on $\mathbb{A}$ of positions labeled i is exactly w_i for $i = 1, 2$.

Synchronized pairs of words (over a fixed alphabet $\mathbb{A}$)

Synchronizing pairs of words

A **synchronization** of (w_1, w_2) is a word over $\mathbf{2} \times \mathbb{A}$ so that the projection on $\mathbb{A}$ of positions labeled i is exactly w_i for $i = 1, 2$.

Example

$(1, a)(1, b)(2, a)$ and $(1, a)(2, a)(1, b)$ synchronize (ab, a).

Synchronized pairs of words (over a fixed alphabet $\mathbb{A}$)

Synchronizing pairs of words

A **synchronization** of (w_1, w_2) is a word over $\mathbf{2} \times \mathbb{A}$ so that the projection on $\mathbb{A}$ of positions labeled i is exactly w_i for $i = 1, 2$.

Example

$(1, a)(1, b)(2, a)$ and $(1, a)(2, a)(1, b)$ synchronize (ab, a).

Every word $w \in (\mathbf{2} \times \mathbb{A})^*$ is a synchronization of a unique pair (w_1, w_2) that we denote $[\![w]\!]$.

$$[\![(1, a)(1, b)(2, a)]\!] = [\![(1, a)(2, a)(1, b)]\!] = (ab, a).$$

Synchronized relations

Synchronizing relations

We lift this notion to languages $L \subseteq (\mathbf{2} \times \mathbb{A})^*$

$$[\![L]\!] = \{[\![w]\!] \mid w \in L\}$$

Example

$\mathbb{A} = \{a, b\}$, $L = ((1,a)(2,a) \cup (1,a)(2,b) \cup (1,b)(2,a) \cup (1,b)(2,b))^*$,

$$[\![L]\!] = \{(w_1, w_2) \mid |w_1| = |w_2|\}.$$

C-controlled relations

Restrictions on the shape of the projection over **2**

$\rightsquigarrow$

Infinitely many different classes of relations.

C-controlled relations

Restrictions on the shape of the projection over $\mathbf{2}$

$\rightsquigarrow$

Infinitely many different classes of relations.

C-controlled words and languages

$C \subseteq \mathbf{2}^*$ regular

- $w \in (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if its projection over $\mathbf{2}$ belongs to C.
- $L \subseteq (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if all its words are.

C-controlled relations

Restrictions on the shape of the projection over $\mathbf{2}$

$\rightsquigarrow$

Infinitely many different classes of relations.

C-controlled words and languages

$C \subseteq \mathbf{2}^*$ regular

- $w \in (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if its projection over $\mathbf{2}$ belongs to C.
- $L \subseteq (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if all its words are.

Examples

Every $w \in (\mathbf{2} \times \mathbb{A})^*$ is $\mathbf{2}^*$-controlled,
$(1, a)(1, b)(2, a)$ is 1^*2^*-controlled,
$(1, a)(2, a)(1, b)$ isn't 1^*2^*-controlled,
L (previous slide) is $(12)^*$-controlled.

C-controlled relations

Restrictions on the shape of the projection over $\mathbf{2}$

⇝

Infinitely many different classes of relations.

C-controlled words and languages

$C \subseteq \mathbf{2}^*$ regular

- $w \in (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if its projection over $\mathbf{2}$ belongs to C.
- $L \subseteq (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if all its words are.

Examples

Every $w \in (\mathbf{2} \times \mathbb{A})^*$ is $\mathbf{2}^*$-controlled,
$(1, a)(1, b)(2, a)$ is 1^*2^*-controlled,
$(1, a)(2, a)(1, b)$ isn't 1^*2^*-controlled,
L (previous slide) is $(12)^*$-controlled.

C-controlled relations

Given a regular language $C \subseteq \mathbf{2}^*$

$$\textsc{Rel}(C) = \{ [\![L]\!] \mid L \text{ is reg. and } C\text{-controlled} \}$$

C-controlled relations

Restrictions on the shape of the projection over $\mathbf{2}$

$\rightsquigarrow$

Infinitely many different classes of relations.

C-controlled words and languages

$C \subseteq \mathbf{2}^*$ regular

- $w \in (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if its projection over $\mathbf{2}$ belongs to C.
- $L \subseteq (\mathbf{2} \times \mathbb{A})^*$ is C-**controlled** if all its words are.

Examples

Every $w \in (\mathbf{2} \times \mathbb{A})^*$ is $\mathbf{2}^*$-controlled,
$(1, a)(1, b)(2, a)$ is 1^*2^*-controlled,
$(1, a)(2, a)(1, b)$ isn't 1^*2^*-controlled,
L (previous slide) is $(12)^*$-controlled.

C-controlled relations

Given a regular language $C \subseteq \mathbf{2}^*$

$$\textsc{Rel}(C) = \{[\![L]\!] \mid L \text{ is reg. and } C\text{-controlled}\}$$

Examples

$\textsc{Rel}(1^*2^*) = \text{REC}$,
$\textsc{Rel}((12)^*(1^* \cup 2^*)) = \text{REG}$,
$\textsc{Rel}(\mathbf{2}^*) = \text{RAT}$.

Class Containment Problem

CLASS CONTAINMENT PROBLEM	
Input:	Two regular languages $C, D \subseteq \mathbf{2}^*$
Output:	Is $\text{REL}(C) \subseteq \text{REL}(D)$?

Examples

If $C \subseteq D$, then $\text{REL}(C) \subseteq \text{REL}(D)$,
$\text{REL}(1^*2^*) \subseteq \text{REL}((12)^*(1^* \cup 2^*))$,
$\text{REL}((12)^*(1^* \cup 2^*)) \not\subseteq \text{REL}(1^*2^*)$,
$\text{REL}(1^*2^*) = \text{REL}(2^*1^*)$,
$\text{REL}((12)^*) = \text{REL}((21)^*)$.

Previous work[2]

Decidability and complexity

The problem is decidable for $\textsc{Rel}(D)$ =REC, REG or RAT.

Resynchronization

The proof is constructive in terms of the automaton:

Given a NFA for a C-controlled language L, one can effectively construct a NFA for a D-controlled language L' such that $[\![L]\!] = [\![L']\!]$.

[2] D. Figueira and L. Libkin. Synchronizing relations on words. *ACM Transactions on Computer Systems, 2015.*

Our contribution

We prove that the Class Containment Problem is decidable for arbitrary C and D and, in case of positive answer, we give an effective method for resynchronizing relations.

Proof idea

Step 1: Rewrite C and D as finite unions of *simple languages*.

Step 2: Characterization for simple languages.

Step 3: Induction on the amount of disjuncts in the unions.

Step 1: Decomposition into simple languages

Concat-star languages

$$C_1^* u_1 \cdots C_n^* u_n$$

with $C_1, \ldots, C_n$ regular languages, $u_1, \ldots, u_n$ words.
A component C_i^* is **homogeneous** if it is contained in 1^* or 2^*.
Otherwise is **heterogeneous**.

- **heterogeneous** if it contains at least one heterogeneous component, otherwise it is **homogeneous**;
- **smooth** if every homogeneous component is 1^{k*} or 2^{k*}, for some $k > 0$, and there are no consecutive homogeneous components;
- **simple** if it has star-height 1 and it is either homogeneous or smooth heterogeneous.

	homogeneous	smooth heterogeneous	non-smooth heterogeneous	non concat-star
s.-h. > 1	$(1^*1)^*2^*$	$1^*(1^*2)^*2^*$	$1^*2^*(1^*2)^*$	$(1^*2)^* \cup (12)^*$
s.-h. = 1	$1^*(11)^*2^*$	$1^*(12)^*2^*$	$1^*2^*(12)^*$	$(12)^*1^* \cup (12)^*2^*$

simple

Step 1: Decomposition into simple languages

Every regular language is a finite union of concat-star languages.

Step 1: Decomposition into simple languages

Every regular language is a finite union of concat-star languages.

Every concat-star language is *Rel-equivalent* to a finite union of concat-star languages of star-height 1.

Step 1: Decomposition into simple languages

Every regular language is a finite union of concat-star languages.

Every concat-star language is *Rel-equivalent* to a finite union of concat-star languages of star-height 1.

Every concat-star language of star-height 1 is *Rel-equivalent* to a finite union of simple languages.

Step 1: Decomposition into simple languages

Every regular language is a finite union of concat-star languages.

Every concat-star language is *Rel-equivalent* to a finite union of concat-star languages of star-height 1.

Every concat-star language of star-height 1 is *Rel-equivalent* to a finite union of simple languages.

Example

$\textsc{Rel}((12)^*1^*2^*) = \textsc{Rel}((12)^*1^* \cup (12)^*2^*)$.

Step 2: Characterization for simple languages

Parikh ratio

$w \in \mathbf{2}^* \setminus \{\varepsilon\}$, $\rho(w) = \frac{|w|_1}{|w|}$.
$C \subseteq \mathbf{2}^*$, $\rho(C) = \{\rho(w) \mid w \in C \setminus \{\varepsilon\}\} \subseteq [0,1]_{\mathbb{Q}}$.

Step 2: Characterization for simple languages

Parikh ratio

$w \in \mathbf{2}^* \setminus \{\varepsilon\}$, $\rho(w) = \frac{|w|_1}{|w|}$.
$C \subseteq \mathbf{2}^*$, $\rho(C) = \{\rho(w) \mid w \in C \setminus \{\varepsilon\}\} \subseteq [0,1]_{\mathbb{Q}}$.

Synchronizing morphisms

$C = C_1^* u_1 \cdots C_n^* u_n$, $D = D_1^* v_1 \cdots D_m^* v_m$. $C \xrightarrow{s.m.} D$ is

$$f : [1, \ldots, n] \to [1, \ldots, m] \text{ s.t.}$$

i) f is monotonic and
ii) $\rho(C_i^*) \subseteq \rho(D_{f(i)}^*)$ for all $i = 1, \ldots, n$.
If C is homogeneous, we have a s.m. to any D by convention.

Step 2: Characterization for simple languages

Parikh ratio

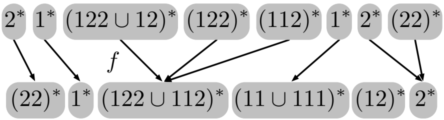
$w \in \mathbf{2}^* \setminus \{\varepsilon\}$, $\rho(w) = \frac{|w|_1}{|w|}$.
$C \subseteq \mathbf{2}^*$, $\rho(C) = \{\rho(w) \mid w \in C \setminus \{\varepsilon\}\} \subseteq [0,1]_{\mathbb{Q}}$.

Synchronizing morphisms

$C = C_1^* u_1 \cdots C_n^* u_n$, $D = D_1^* v_1 \cdots D_m^* v_m$. $C \xrightarrow{s.m.} D$ is

$$f : [1, \ldots, n] \to [1, \ldots, m] \text{ s.t.}$$

i) f is monotonic and
ii) $\rho(C_i^*) \subseteq \rho(D_{f(i)}^*)$ for all $i = 1, \ldots, n$.
If C is homogeneous, we have a s.m. to any D by convention.



Step 2: Characterization for simple languages

Proposition

For all simple languages $C, D \subseteq \mathbf{2}^*$, $\textsc{Rel}(C) \subseteq \textsc{Rel}(D)$ iff $\pi(C) \subseteq \pi(D)$ and $C \xrightarrow{s.m.} D$.

Step 2: Characterization for simple languages

Proposition

For all simple languages $C, D \subseteq \mathbf{2}^*$, $\textsc{Rel}(C) \subseteq \textsc{Rel}(D)$ iff $\pi(C) \subseteq \pi(D)$ and $C \xrightarrow{s.m.} D$.

Examples

$\textsc{Rel}((12)^*(112)^*) \subseteq \textsc{Rel}((12 \cup 11122)^*(121)^*1^*2^*)$,
$\textsc{Rel}((112)^*(12)^*) \not\subseteq \textsc{Rel}((12 \cup 11122)^*(121)^*1^*2^*)$.

Step 3: Dealing with unions

Unions on the left

$\textsc{Rel}(C_1 \cup C_2) \subseteq \textsc{Rel}(D)$ iff
$\textsc{Rel}(C_1) \subseteq \textsc{Rel}(D)$ and $\textsc{Rel}(C_2) \subseteq \textsc{Rel}(D)$.

Main theorem

For C simple and $D = \bigcup_j D_j$ a finite union of simple languages, the following are equivalent:

i) $\textsc{Rel}(C) \subseteq \textsc{Rel}(D)$,

ii) $\pi(C) \subseteq \pi(D)$, $\exists j$ with $C \xrightarrow{s.m.} D_j$ and in addition, if C is heterogeneous, then $\textsc{Rel}(C \setminus [D_j]_\pi) \subseteq \textsc{Rel}(\bigcup_{j' \neq j} D_{j'})$.

Future work

- Our proof gives an effective algorithm to resynchronize relations. We would like to determine the exact complexity.

- Other natural questions about this framework: existence and computability of canonical control languages, for which control languages C $\textsc{Rel}(C)$ is closed under intersection, etc.

Thanks for your attention!