Word Languages from Trace Viewpoint

Barbara Klunder, Krystyna Stawikowska

UMK Toruń
presented by Łukasz Mikulski

October 2008

Traces (Mazurkiewicz 1977)

Concurrent alphabet (A, I) or (A, D)

Independency $I \subseteq A \times A$ is symmetric and irreflexive
Dependency $D \subseteq A \times A$ is a complement of I.

Equivalence of words

The congruence $\equiv \subseteq A^\star \times A^\star$ is generated by the set of all pairs $ab \equiv ba$ where aIb.

Trace monoid

Trace monoid is a quotient $A^*/_\equiv$, denoted in the sequel by $A^\star/I$.
Elements of $A^\star/I$ are called traces, subsets are called trace languages.

Flat and trace languages

Trace languages generated by flat languages

Let $w \in A^\star$ and $[w] \in A^\star/I$ be the congruence class of w.

- Flat language $L \subseteq A^\star$ induces a trace language $[L] = \{[w] | w \in L\}$ - the set of all traces induced by members of L.
- Any class $\mathcal{R} \subseteq 2^{A^\star}$ of flat languages induces a class of trace languages $[\mathcal{R}] \subseteq 2^{A^\star/I}$ - the set of all trace languages induced by flat languages of $\mathcal{R}$.

Flattening of trace language

- The flattening of $T \subseteq A^\star/I$ is a flat language $\bigcup T = \{w \in A^\star | [w] \in T\}$.
- A flat language L is said to be *closed* (w.r.t. I) iff $L = \bigcup[L]$.

Flat language $L \subseteq A^\star$

- is *recognizable* if it is recognizable by a finite automaton,
- is *rational or regular* if it is defined by a regular expression.

Let $\mathbf{Reg}_A$ denote the set of all regular languages over A.

Trace language $T \subseteq A^\star/I$ is

- *rational* iff $T = [L]$ for some $L \in \mathbf{Reg}_A$,
- is *recognizable* iff $\bigcup T = \{w \in A^\star | [w] \in T\} \in \mathbf{Reg}_A$.

Let $\mathbf{Rec}(M)$ denote the set of all recognizable subsets of a given monoid M.

Definition (Rank of distribution)

(A, I) - concurrent alphabet, $L \subseteq A^\star$, $u, v \in A^\star$ and $[uv] \in [L]$. We say that $rank_{L,I}(u, v) \leq n$ if there is $w \in L$ such that:

$$w = v_0 u_1 v_1 \ldots u_n v_n, \quad u_1 \ldots u_n \in [u] \quad v_0 v_1 \ldots v_n \in [v]$$
$$\text{and } (v_i, u_j) \in I \quad \text{for all } i < j.$$

And $rank_{L,I}$ is defined as:

$$rank_{L,I}(u, v) = \min\{n \in \mathbb{N} | rank_{L,I}(u, v) \leq n\} \text{ if } [uv] \in [L]$$

The *rank* of a language $L \subseteq A^\star$ (w.r.t. I) is an integer (if it exists) or infinity

$$Rank_I(L) = \begin{cases} \max\{rank_{L,I}(u, v) | [uv] \in [L]\} & \text{if the set is bounded} \\ \infty & \text{otherwise.} \end{cases}$$

FinRank$_I(A)$ - the class of all regular languages of finite rank

$$\mathbf{FinRank}_I(A) = \{L \in \mathbf{Reg}(A) | Rank_I(L) < \infty\}.$$

Lemma (Hashiguchi)

If $L \in$ **FinRank** *then* $[L] \in$ **Rec**.

Countrexample for converse

$A = \{a, b\}$, aIb, $L = (ab)^\star(a^\star \cup b^\star) \subseteq A^\star$

$[L] = A^\star/I \in$ **Rec** and $Rank_I(L) = \infty$, since $rank(a^n, b^n) = n$.

Connected words, traces and languages

Let (A, I) be a concurrent alphabet; let $D = A \times A \setminus I$ be the dependency relation.

- A *word* $w \in A^\star$ is *connected* (w.r.t. D) iff the graph $D|_{Alph(w)}$ is connected;
- A *flat language* $L \subseteq A^\star$ is *connected* iff all its members are connected;
- A *trace* $[w] \in A^\star/I$ is *connected* iff the word w is connected;
- A *trace language* $T \subseteq A^\star/I$ is *connected* iff all its members are connected.

Theorem (Ochmański,Clerbout/Latteux)

Let $A^\star/I$ be a trace monoid.
If $T \subseteq A^\star/I$ is recognizable and connected, then $T^\star$ is recognizable.

Example

Let $(A, I) = (\{a, b\}, \{(a, b), (b, a)\})$, the trace $[ab]$ is not connected. The trace language $T = \{[ab]\}$ is an example of a recognizable language such that $T^\star$ is not recognizable.

Definition (Star-connected expressions and languages)

- A *rational expression* (over A) is *star-connected* iff it has a connected expression under any of its stars.
- A *flat language* $L \subseteq A^\star$ is *star-connected* iff there is a star-connected expression R such that $L = L(R)$.
- $\mathbf{StarCon}_I(A)$ is the class of all star-connected (w.r.t. I) languages (over A).
- A *trace language* $T \subseteq A^\star/I$ is *star-connected* iff there is a star-connected expression R such that $T = [L(R)]$.

Theorem (Ochmański, 1985)

$\mathbf{Rec} = [\mathbf{StarCon}_I(A)]$, *i.e.*

$$T \in \mathbf{Rec} \text{ iff } (\exists L \in \mathbf{StarCon}_I(A))\ T = [L].$$

Example

Any star-connected language has infinitely many rational expressions, defining it. For instance:

$$\begin{aligned} (a \cup b)^\star &= (a^\star b^\star)^\star \\ (a \cup b)^\star b &= (a^\star \cup bb^\star a)^\star bb^\star \end{aligned}$$

Notice that, if aIb, then the left-hand-side expressions are star-connected, whereas the right-hand-side ones are not.

Example

For $X = \{x_1, \dots, x_n\} \subseteq A$ and S_n the set of all permutations of $\{1, \dots n\}$. The star-connected expression

$$R_X = \bigcup_{\sigma \in S_n} X^\star x_{\sigma(1)} X^\star x_{\sigma(2)} \dots X^\star x_{\sigma(n)} X^\star$$

defines the language $L_X = \{w \in A^\star | Alph(w) = X\}$.

Example

Let $\mathcal{C}(A, I) = \{X \subseteq A \mid (X, D|_X) \text{ is connected}\}$.
The star-connected expression

$$C_{(A,I)} = \bigcup_{X \in \mathcal{C}(A,I)} R_X$$

defines the language $Con(A, I)$ of all connected words over the alphabet (A, I).

Lemma

Let $L, K \subseteq A^\star$ have finite Ranks; assume $Rank(L) = n$ and $Rank(K) = m$. Then:

$$Rank(L \cup K) \leq \max(n, k) \tag{1}$$
$$Rank(LK) \leq n + m \tag{2}$$
$$\text{If } L \text{ is connected, then } Rank(L^\star) \leq (n+1)|A| + 1 \tag{3}$$

Example

The language $L = (ab \cup ba \cup aa \cup bb)^\star$, with $a|b$.
$Rank(L) = 1 < \infty$, because $L = \bigcup[L]$ is closed, but L is not star-connected (it will be showed).

Theorem

The class **StarCon** *is properly included in the class* **FinRank**.

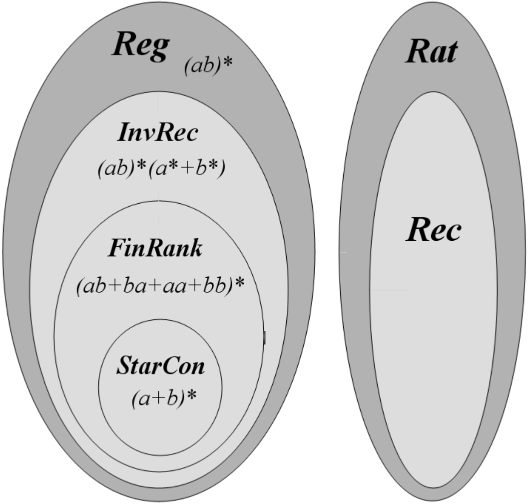

Figure: Classes inducing rational and recognizable trace languages

Question

Are the classes **FinRank** and **StarCon** boolean algebras?

Example (Complement leads out of **FinRank** and **StarCon**)

Let $a I b$. Clearly, $L = (ab)^\star$ is not **FinRank** nor **StarCon**, but its complement

$$A^\star \setminus L = b(a \cup b)^\star \cup (a \cup b)^\star a \cup (a \cup b)^\star aa(a \cup b)^\star \cup (a \cup b)^\star bb(a \cup b)^\star$$

is star-connected.

Example (Intersection leads out of **FinRank**)

Let $a I b$. Let $L = a^\star (ab)^\star b^\star$ and $K = b^\star (ab)^\star a^\star$. One can prove that $Rank(L) = Rank(K) = 2$. But the intersection $L \cap K = (ab)^\star \cup a^\star \cup b^\star$ has infinite Rank.

Lemma (About intersections, Klunder 2005)

Let L be a star-connected flat language and let $X \subseteq A$. Then the flat language $L \cap L_X$ is star-connected. If L is defined by a star-connected expression R then we can effectively find the star-connected expression R_X defining $L \cap L_X$.

Corollary

Let L be a star-connected flat language. Let $X \subseteq A$. Then the flat languages $Con(X, I) = X^\star \cap Con(A, I)$ and $L \cap Con(X, I)$ are star-connected.

Definition (Composition of cycles)

Let $\mathbf{A} =< A, Q, \delta, q, F >$ be an automaton. A *cycle* of $\mathbf{A}$ is a path $s_0 a_1 s_1 a_2 s_2 \ldots a_n s_n$ such that $n > 0$ and $s_0 = s_n$; a cycle is *simple* if $s_i \neq s_j$ whenever $i \neq j$ (for $i, j = 0, \ldots, n-1$). A cycle is *connected* if its label $w = a_1 \ldots a_n$ is a connected word. If for some $k \geq 0$ and $i_0 = 0 < i_1 < \ldots < i_k < i_{k+1} = n$ equations $s_{i_j} = s_0$ hold then the cycle is a *composition* in s_0 of cycles $s_{i_j} a_{i_j+1} \ldots s_{i_{j+1}}$ for $0 \leq j \leq k$. If the sequence $i_0 = 0 < i_1 < \ldots < i_k < i_{k+1} = n$ contains all occurrences of s_0 then the composition is proper.

Isomorphism of cycles

For every $0 < i < n$ the cycle $s_i a_{i+1} \ldots a_n s_n a_1 s_1 \ldots a_i s_i$ is *isomorphic* to a cycle $s_0 a_1 s_1 a_2 s_2 \ldots a_n s_n$ of a given automaton $\mathbf{A}$.

Example

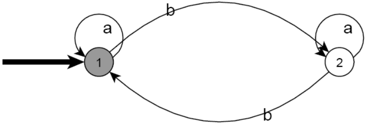

Figure: Automaton with connected simple cycles

Let $a I b$. Then every simple cycle of $\mathbf{A}$ is connected, but the cycle with a label *ababa* starting in initial state 1 is not isomorphic to any cycle which is a proper composition of connected cycles. We can show that the language $L(\mathbf{A})$ of all words with even number occurrences of b is not star-connected.

Necessary condition

Lemma

For every star-connected language $R \subseteq A^+$ there exists an automaton $\mathbf{A} =< A, Q, \delta, q, f >$ with a final state $f \in Q$, such that

1. $R = L(\mathbf{A})$, *i.e.* R *is accepted by* $\mathbf{A}$;
2. *every simple cycle of* $\mathbf{A}$ *is connected;*
3. *for every cycle* $s_0 a_1 s_1 a_2 s_2 \ldots a_n s_n$ *there exist* $0 \leq i \leq n$ *such that the isomorphic cycle* $s_i a_{i+1} \ldots a_n s_n a_1 s_1 \ldots a_i s_i$ *is a proper composition in* s_i *of connected cycles;*
4. *for every cycle* $s_0 a_1 s_1 a_2 s_2 \ldots a_n s_n$ *such that* $s_i = f$ *for some* $0 \leq i \leq n$ *the isomorphic cycle* $s_i a_{i+1} \ldots a_n s_n a_1 s_1 \ldots a_i s_i$ *is a proper composition in* s_i *of connected cycles.*

Sufficient condition

Projection on a subset X of A

$\pi_X : A \mapsto X^\star$ is defined by conditions

$$\pi_X(a) = a \text{ if } a \in X, \pi_X(a) = \varepsilon \text{ otherwise.}$$

The unique extension of π_X to $A^\star$ we call the projection (on X) and denote in the same way.

Proposition

Let (A, I) be any concurrent alphabet. Let $X \subseteq A$ be such that X and $A \setminus X$ are independent: $X \times (A \setminus X) \subseteq I$. Let $\mathbf{A} =< A, Q, \delta, q, F >$ be any automaton satisfying conditions 2 and 3 of the previous lemma. Then the language $\pi_X(L(\mathbf{A}))$ is star-connected.

Crucial step of the inductive proof

For $S \subseteq Q$ and $i, j \in Q$ let L_{ij}^{S} be the set of all projections of labels of paths of **A** starting in i, ending in j and going through the states of S. Then

$$L_{ij}^{S} = \bigcup_{k \in S} L_{ik}^{S\setminus\{k\}} (L_{kk}^{S\setminus\{k\}})^{\star} L_{kj}^{S\setminus\{k\}} \cup \bigcup_{k \in S} L_{ij}^{S\setminus\{k\}}$$

But we claim that

$$(L_{kk}^{S\setminus\{k\}})^{\star} = \bigcup_{j \in S} L_{kj}^{S\setminus\{j\}} (L_{jj}^{S\setminus\{j\}} \cap Con(X, I))^{\star} L_{jk}^{S\setminus\{j\}}$$

so this language is star-connected.

New operations on star-connected flat languages

Lemma

Let L be star-connected flat language. Then the following languages are star-connected:

1. *the decomposition of L:*

$$/L/ = \bigcup_{\{X \in \mathcal{C}(A,I) | X \times (A \setminus X) \subseteq I\}} \pi_X(L) \cap L_X$$

2. *the concurrent iteration of L:* $L^{\otimes} = /L/^{\star}$.

If L is defined by a star-connected expression R then we can effectively find star-connected expressions R_d, R_c *such that* $L(R_d) = /L/$ *and* $L(R_c) = L^{\otimes}$.

Summary

Theorem (About characterization)

Let (A, I) be any concurrent alphabet. Then the language $L \subseteq A^\star$ is star-connected iff L is accepted by some automaton satisfying the following conditions:

1. *every simple cycle of* $\mathbf{A}$ *is connected;*
2. *for every cycle $s_0 a_1 s_1 a_2 s_2 \ldots a_n s_n$ there exist $0 \leq i \leq n$ such that the isomorphic cycle $s_i a_{i+1} \ldots a_n s_n a_1 s_1 \ldots a_i s_i$ is a proper composition in s_i of connected cycles.*

Corollary

Let $L \in \mathbf{StarCon}_I(A)$ and $M \subseteq A^\star$. Then the language, quotient of a star-connected language, $L/M = \{w | \exists_{v \in M} wv \in L\}$ is star-connected.

Question

Is the property characterizing automata accepting star-connected languages preserved in the process of determinization or minimalization?

Positive answer implies

the decidability of the following problem:

Is a regular language L star-connected?

Example

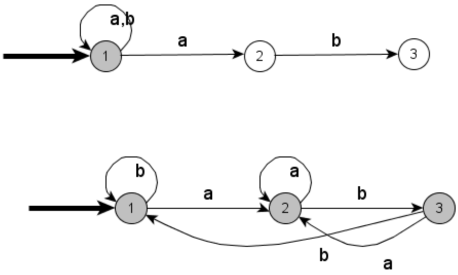

Figure: Deterministic automaton without connected simple cycles

Let *a*|*b*. The second automaton (without connected cycles) is the result of determinization of the first one.

Example

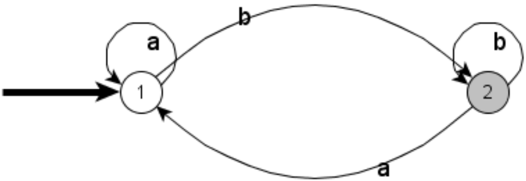

Figure: Minimal deterministic automaton without connected simple cycles

Let aIb. Then the simple cycle $1b2a1$ of **A** is not connected, but the language accepted by **A** of all words ending with b is defined by the following star-connected expression: $(a \cup b)^\star b$.

Lemma (Pumping Lemma)

Let $L \subseteq A^\star$ be a star-connected language over a concurrent alphabet (A, I).

1. *There exists $n > 0$ such that, for all $z \in L$ if $|z| \geq n$ then for some $u, v, w \in A^\star : z = uvw$, v is nonempty and connected, $|uv| \leq n$ and $uv^iw \in L$ for all $i \in N$.*
2. *If the dependency relation $D = A \times A \setminus I$ is transitive, then there exist $m > 0$ such that for all $z \in L$ if $|z| \geq m$ then for some $u, v_1, v_2, w \in A^\star$: $z = uv_1v_2w$, words v_1, v_2 are nonempty and connected, $u(v_1 \cup v_2)^\star w \subseteq L$.*

Example

Let aIb and $L = (aa \cup ab \cup ba \cup bb)^*$. Using Pumping Lemma, one can easly show that the languages L is not star-connected.

Open problems

1. The status of the decision problem "Is a regular language L star-connected?" is unknown.
2. Is the class of star-connected flat languages closed under intersection?

Sources of the talk

1. B. Klunder: Characterization of Star-Connected Languages using Finite Automata. Proceedings of LATA 2008, pp. 325-334. Universitat Tarragona, Spain, 2008;
2. B. Klunder, E. Ochmański, K. Stawikowska: On Star-Connected Flat Languages. Fundamenta Informaticae 67(1-3), pp. 93-105, 2005;
3. K. Stawikowska: Word Languages Inducing Recognizable and Star-Free Trace Languages. PhD Thesis, Toruń/Warsaw, 2007.