Forbidden Patterns for Ordered Automata

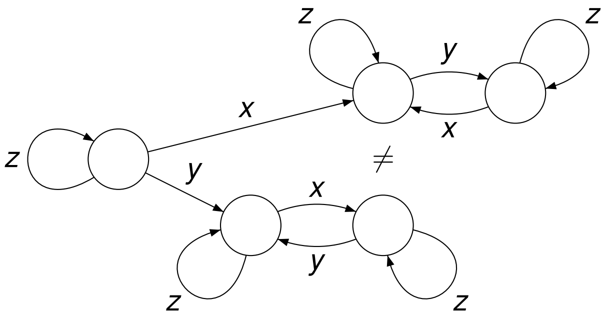
Ondřej Klíma and Libor Polák

Department of Mathematics and Statistics
Masaryk University, Brno
Czech Republic

Košice, NCMA 2018

Example

Cohen, Perrin and Pin (1993): a language L is expressible by the restriction of linear temporal logic obtained by considering only the operators "next" and "eventually" if and only if the minimal automaton of L does not contain the following pattern.



Our Aim

- We would like to develop a unified theory of forbidden patterns which would explain some general behaviour of these patterns and the classes defined by them and compare formalisms in numerous known applications.

Our Aim

- We would like to develop a unified theory of forbidden patterns which would explain some general behaviour of these patterns and the classes defined by them and compare formalisms in numerous known applications.
- In particular, the considered classes of languages are closed with respect to natural operations.

Our Aim

- We would like to develop a unified theory of forbidden patterns which would explain some general behaviour of these patterns and the classes defined by them and compare formalisms in numerous known applications.
- In particular, the considered classes of languages are closed with respect to natural operations.
- We develop our theory in the framework of the theory of varieties of regular languages.
(More generally: positive varieties or $\mathcal{C}$-varieties.)

Notation and Terminology

- In this presentation, automata are finite and deterministic. We considered only regular languages.
- A semiautomaton is a deterministic automaton without initial and final states being specified.
- A (semi)automaton is always complete. If it is not the case, we talk about a partial (semi)automaton.

Notation and Terminology

- In this presentation, automata are finite and deterministic. We considered only regular languages.
- A semiautomaton is a deterministic automaton without initial and final states being specified.
- A (semi)automaton is always complete. If it is not the case, we talk about a partial (semi)automaton.
- More formally: A partial semiautomaton is a triple $(Q, A, \cdot)$, where Q is a fintie set of states, A is an alphabet and $\cdot : Q \times A \to Q$ is a partial transition function.
- A homomorphism of a partial semiautomaton $(P, A, \cdot)$ to a partial semiautomaton $(Q, A, \cdot)$ is a mapping $\varphi : P \to Q$ such that: for each $a \in A$ and $p, q \in P$ satisfying $p \cdot a = q$, we have $\varphi(p) \cdot a = \varphi(q)$.

Varieties of Regular Languages

Definition

A variety of languages $\mathcal{V}$ associates to every finite non-empty alphabet A a class $\mathcal{V}(A)$ of regular languages over A in such a way that

- $\mathcal{V}(A)$ is closed under finite unions, finite intersections and complements ($\emptyset, A^* \in \mathcal{V}(A)$),
- $\mathcal{V}(A)$ is closed under quotients, i.e.
  $L \in \mathcal{V}(A),\ u, v \in A^*$ implies
  $u^{-1}Lv^{-1} = \{\, w \in A^* \mid uwv \in L \,\} \in \mathcal{V}(A)$,
- $\mathcal{V}$ is closed under preimages in morphisms, i.e.
  $f : B^* \to A^*,\ L \in \mathcal{V}(A)$ implies
  $f^{-1}(L) = \{\, v \in B^* \mid f(v) \in L \,\} \in \mathcal{V}(B)$.

Eilenberg Correspondence

- Eilenberg correspondence: one-to-one correspondence between varieties of regular languages and pseudovarieties of monoids.
- A pseudovariety of finite monoids is a class of finite monoids closed under submonoids, homomorphic images and products of finite families.

Eilenberg Correspondence

- Eilenberg correspondence: one-to-one correspondence between varieties of regular languages and pseudovarieties of monoids.
- A pseudovariety of finite monoids is a class of finite monoids closed under submonoids, homomorphic images and products of finite families.
- Well established modifications:
  - positive varieties (Pin) – classes need not to be closed under complementation,
  - $\mathcal{C}$-varieties (Straubing) – classes are closed only under preimages in morphisms from a fixed family of morphism $\mathcal{C}$,
  - positive $\mathcal{C}$-varieties – a common generalization.

Motivation by a Famous Example

Simon's characterization of piecewise testable languages:
Piecewise testable languages correspond to $\mathcal{J}$-trivial monoids.

- We want to recognize the piecewise testability from the (minimal) automaton.
- In the original paper, Simon gave two conditions which can be checked in polynomial time as shown by Stern latter.
- What kind of classes of automata corresponds to varieties of languages?

Motivation by a Famous Example

Simon's characterization of piecewise testable languages:
Piecewise testable languages correspond to $\mathcal{J}$-trivial monoids.

- We want to recognize the piecewise testability from the (minimal) automaton.
- In the original paper, Simon gave two conditions which can be checked in polynomial time as shown by Stern latter.
- What kind of classes of automata corresponds to varieties of languages?
- Matching operations:
  - quotients . . . changing initial and final states
  - intersection of languages . . . product of automata
  - union of languages . . . product of automata
  - complement . . . changing final states
  - preimage in a morphism f . . . f-renaming of automata

Varieties of Semiautomata

We add also operations on automata which does not extend the classes of recognizable languages, e.g., homomorphic images.

Definition

A variety of semiautomata $\mathbb{V}$ associates to every non-empty finite alphabet A a class $\mathbb{V}(A)$ of semiautomata over alphabet A in such a way that

- a one-element semiautomaton over A is in $\mathbb{V}(A)$ and this class is closed under disjoint unions and direct products of pairs, subsemiautomata and homomorphic images,
- $\mathbb{V}$ is closed under renaming.

Eilenberg type correspondence was given by Chaubard, Pin and Straubing (semiautomata are called actions).

Partial Order on Minimal Automata

- The modification to $\mathcal{C}$-varieties is straightforward.
  (It is not a purpose of the talk, it is explained in the paper).
- The modification to positive varieties uses

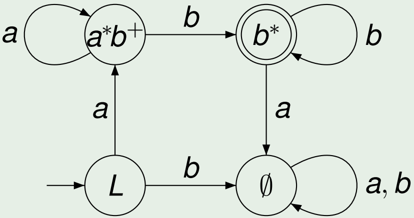
Partial Order on Minimal Automata

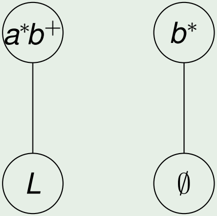
- The modification to $\mathcal{C}$-varieties is straightforward.
  (It is not a purpose of the talk, it is explained in the paper).
- The modification to positive varieties uses ordered structures (monoids, automata).
- We explain that the minimal automaton of a given language is implicitly ordered:
  - One can assign to each state q its *future* consisting of all words which are acceptable if q would be the initial state.
  - Different states have different futures (from the minimality).
  - Now, if we identify states with their futures, then the relation $\subseteq$ is a partial order on the set of states.
  - It is compatible with every action by a single letter.
  - The final states form upward closed subset.

Example

Example

- The language $L = a^+b^+$ is piecewise testable:
  $L = A^*aA^*bA^* \cap (A^*bA^*aA^*)^c$.
- The minimal automaton: The partial order:





- Non-final states do not form an upward closed subset.

The Notion of Ordered Semiautomata

Definition

An ordered semiautomaton is $\mathcal{A} = (Q, A, \cdot, \leq)$, where $(Q, A, \cdot)$ is a semiautomaton, $(Q, \leq)$ is an ordered set and for every pair of states $p \leq q$ and a letter $a \in A$ we have $p \cdot a \leq q \cdot a$.

The transition function can be extended to a mapping $\cdot : Q \times A^* \to Q$ in a usual way.

Definition

The ordred semiautomaton $\mathcal{A}$ recognizes a language L if there is a state $i \in Q$ and upward closed subset F of Q such that

$$L = \{u \in A^* \mid i \cdot u \in F\} .$$

There is a one-to-one correspondence between $\mathcal{C}$-varieties of ordered semiautomata and positive $\mathcal{C}$-varieties of languages.

Satisfying Configurations

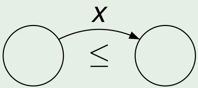
Definition

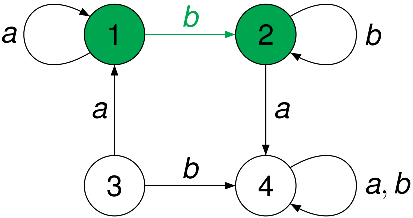
- A configuration $\mathcal{K} = (\mathcal{G}, k, l)$ consists of a partial semiautomaton $\mathcal{G} = (V, X, \cdot)$ and a pair of states $k, l \in V$.
- Let a partial semiautomaton $\mathcal{G} = (V, X, \cdot)$ and an ordered semiautomaton $\mathcal{A} = (Q, A, \cdot, \leq)$ be given. We say that a mappings $\varphi : V \to Q$ and $g : X \to A^*$ are compatible if, for every $m \in V$ and $x \in X$ such that $m \cdot x$ is defined in $\mathcal{G}$, we have $\varphi(m) \cdot g(x) = \varphi(m \cdot x)$.
- We say that an ordered semiautomaton $\mathcal{A} = (Q, A, \cdot, \leq)$ satisfies the configuration $\mathcal{K} = (\mathcal{G}, k, l)$ if, for every pair of compatible mappings $\varphi : V \to Q$ and $g : X \to A^*$ we have that $\varphi(k) \leq \varphi(l)$.

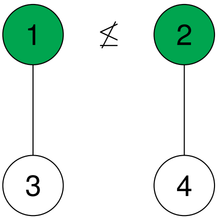
Example

Example

A regular language is a positive piecewise testable if and only if its minimal ordered (semi)automaton satisfies the following configuration:







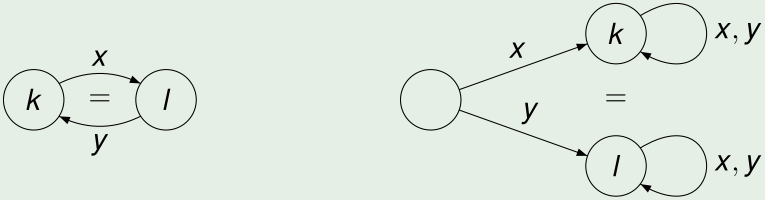
$L = a^+b^+$ is not a positive piecewise testable language.

Comments

- In the paper, a more general case of the definitions is considered, namely g's are taken only from a class of substitutions which is a part of the configuration.
- In the unordered case, we have $\varphi(k) = \varphi(l)$ instead of $\varphi(k) \leq \varphi(l)$.

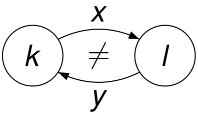
Example

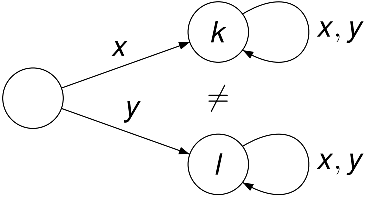
The configurations for two Simon's conditions are the following:



Forbidden Patterns I

- In the paper, we survey numerous known examples of forbidden patterns.
- The first type of forbidden patterns (in the unordered case): one only replaces $=$ by $\neq$ in a configuration and it is required to avoid the pattern.
- In the literature, the patterns are often viewed as subautomata, which leads to inaccuracies.



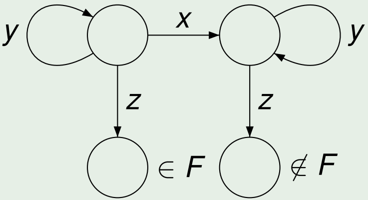


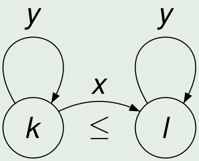
Forbidden Patterns II

- The second type of forbidden patterns: a pattern is enriched by a condition that a certain state is final and another one is non-final. (The corresponding class of languages is not closed under complements.)
- Under a special assumption we were able to give the equivalent configuration/pattern using ordered automata.

Example (The level 1/2 in the dot-depth)

The forbidden pattern and the configuration.

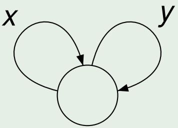




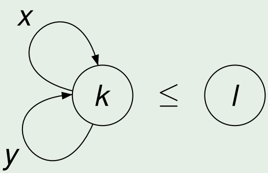
Forbidden Patterns III

- The third type of forbidden patterns: other special conditions are required.
- Example: sparse languages (those of polynomial density).
- Here words which are substituted for x and y need to have a different first letters.
- The sparse languages form a positive $\mathcal{C}$-variety for a special $\mathcal{C}$, therefore the configuration works only with special substitutions.

Example (Sparse languages)



$i(x) \neq i(y)$



Basic Properties of Configurations

We denote by $\mathbb{K}(A)$ the class of all ordered semiautomata over A satisfying the configuration $\mathcal{K}$.

Proposition

Let $\mathcal{K} = (\mathcal{G}, k, l)$ be an arbitrary configuration and let A be a non-empty finite set. Then the following hold.

- $\mathbb{K}(A)$ *contains the one-element semiautomaton.*
- *If $\mathcal{G}$ is connected, then $\mathbb{K}(A)$ is closed with respect to disjoint unions.*
- $\mathbb{K}(A)$ *is closed with respect to subsemiautomata.*
- $\mathbb{K}(A)$ *is closed with respect to products of pairs.*
- $\mathbb{K}$ *is closed with respect to renaming.*

Problem: homomorphic images.

Basic Result

Theorem

Let $\mathcal{K} = (\mathcal{G}, k, l)$ be a configuration such that $\mathcal{G}$ is connected, balanced and simple partial semiautomaton. Then the class of all ordered semiautomata satisfying $\mathcal{K}$ forms a variety of ordered semiautomata.

In the paper, you can also find:

- a more general version of the theorem for $\mathcal{C}$-varieties of ordered semiautomata.
- the survey on known examples: reversible languages (two variants), locally $\mathcal{R}$-trivial semigroups, low levels in the concatenation hierarchies, sparse languages.

Thank you.