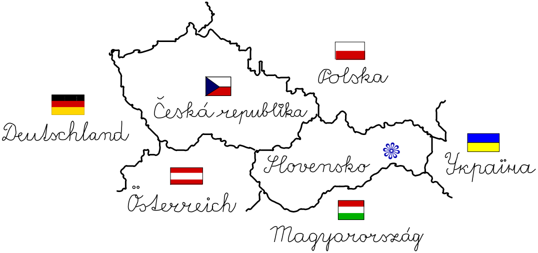
Star-Complement-Star and Kuratowski Algebras on Prefix-Free Languages

Galina Jirásková and Matúš Palmovský

Mathematical Institute, Slovak Academy of Sciences, Košice, Slovakia



Outline

1. Kuratowski 14-theorem in the settings of formal languages
2. Star-complement-star
   - on regular languages
   - on prefix-free regular languages
3. Kuratowski algebras generated by
   - prefix-free languages
   - factor-free and subword-free languages

Kuratowski 14-theorem in the settings of formal languages

- repeatedly apply the operations of closure and complement

Theorem (Kuratowski 1922)

In any topological space: $\leq$ *14 distinct sets can be produced*

Theorem (Brzozowski, Grant, Shallit 2011)

- *Positive closure and complement:* $\leqslant$ *10 distinct languages*
- *Kleene closure and complement:* $\leqslant$ *14 distinct languages* $L, L^*, L^{c*}, L^{*c*}, L^{c*c*}$ *and their complements (up to* ε*)*

- a motivation for the star-complement-star operation
- is its complexity double-exponential?

Star-complement-star on regular languages

Theorem (Jiraskova, Shallit 2012)

The state complexity star-complement-star is $2^{\Theta(n \log n)}$.

- more precisely:

$$2^{\frac{1}{8} n \log n} \leq \cdot \leq 2^{3n \log n}$$

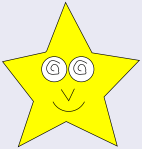
Star-complement-star on regular languages

Theorem (Jiraskova, Shallit 2012)

The state complexity star-complement-star is $2^{\Theta(n \log n)}$.

- more precisely:

$$2^{\frac{1}{8} n \log n} \leq \cdot \leq 2^{3n \log n}$$



- star is an easy operation on prefix-free languages
- n-state (minimal) DFA for prefix-free $L \Rightarrow$ n-state DFA for L^*

- can we get the exact complexity of L^{*c*} on prefix-free languages?

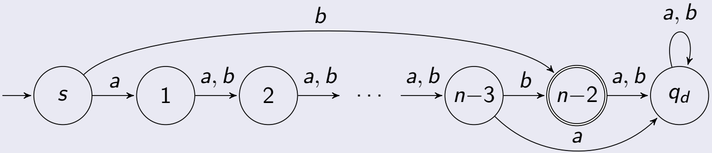
Star-complement-star on prefix-free languages

Theorem

The complexity of star-complement-star on prefix-free languages is

$$f(n) = \begin{cases} 1, & \text{if } n = 1 \text{ or } n = 2; \\ 2^{n-3} + 2, & \text{if } n \geq 3. \end{cases}$$

Binary witness language:



Unary case: always 1, except for $n = 2$ where it is 2

Kuratowski algebras: Positive closure and complement

Theorem (Brzozowski, Grant, Shallit 2011)

- *Start with any language L, and apply the operators of positive closure and complement in Σ^* in any order, any number of times. Then at most 10 distinct languages are generated, and this bound is optimal.*
- *Furthermore, the next table describes the 9 algebras generated by this process, classifies languages according to the algebra they generate, and gives a language generating each algebra.*

Case	Necessary and Sufficient Conditions	$\|B(L)\|$	$\|A(L)\|$	Example
(1)	L is clopen.	1	2	a^*
(2)	L is open but not closed.	2	4	a
(3)	L is closed but not open.	2	4	aaa^*
(4)	L is neither open nor closed; L^+ is clopen and $L^{\oplus +} = L^+$.	3	6	$a \cup aaa$
(5)	L is neither open nor closed; $L^\oplus$ is clopen and $L^{+\oplus} = L^\oplus$.	3	6	aa
(6)	L is neither open nor closed; L^+ is open but $L^\oplus$ is not closed; $L^{\oplus +} \neq L^+$.	4	8	$G := a \cup abaa$
(7)	L is neither open nor closed; $L^\oplus$ is closed but L^+ is not open; $L^{+\oplus} \neq L^\oplus$.	4	8	$(a \cup b)^+ \setminus G$
(8)	L is neither open nor closed; $L^\oplus$ is not closed and L^+ is not open; $L^{+\oplus} = L^{\oplus +}$.	4	8	$a \cup bb$
(9)	L is neither open nor closed; $L^\oplus$ is not closed and L^+ is not open; $L^{+\oplus} \neq L^{\oplus +}$.	5	10	$a \cup ab \cup bb$

BGS 2011: Classification of formal languages by the structure of algebra $(B(L),^+,^\oplus)$.

Interior. Closed, open, and clopen languages.

Definition

The (positive) interior of L is $L^{\oplus} = L^{c+c}$ (we have $L^{\oplus} \subseteq L \subseteq L^{+}$).
A language L is

- closed if $L = L^{+}$
- open if L^{c} is closed (equivalently if $L = L^{\oplus}$)
- clopen if it is both closed and open

Example

(1) closed: $\emptyset, \{\varepsilon\}, \Sigma^{+}, \Sigma^{*}$
(2) open: $\Sigma^{*}, \Sigma^{+}, \{\varepsilon\}, \emptyset$ $\Rightarrow$ they are clopen
(3) $\{a\}$ is not closed but it is open $\{a\}^{\oplus} = \{a\}$
(4) $\{a^{n-2}\}$ is neither closed nor open if $n \geq 4$ $a^{n-2} = a \cdot a^{n-3}$
(5) If L is a prefix-free and clopen, then $L = \emptyset$ or $L = \{\varepsilon\}$

Families $A(L)$ and $B(L)$

$A(L)$ = the family of languages generated from L by positive closure and complement

$B(L)$ = the family of languages generated from L by positive closure and positive interior

Lemma (Brzozowski, Grant, Shallit 2011)

$$A(L) = B(L) \,\dot\cup\, \{M^c \mid M \in B(L)\}$$

Case	Necessary and Sufficient Conditions	$\vert B(L)\vert$	$\vert A(L)\vert$	Example
(1)	L is clopen.	1	2	a^*
(2)	L is open but not closed.	2	4	a
(3)	L is closed but not open.	2	4	aaa^*
(4)	L is neither open nor closed; L^+ is clopen and $L^{\oplus+} = L^+$.	3	6	$a \cup aaa$
(5)	L is neither open nor closed; $L^\oplus$ is clopen and $L^{+\oplus} = L^\oplus$.	3	6	aa
(6)	L is neither open nor closed; L^+ is open but $L^\oplus$ is not closed; $L^{\oplus+} \neq L^+$.	4	8	$G := a \cup abaa$
(7)	L is neither open nor closed; $L^\oplus$ is closed but L^+ is not open; $L^{+\oplus} \neq L^\oplus$.	4	8	$(a \cup b)^+ \setminus G$
(8)	L is neither open nor closed; $L^\oplus$ is not closed and L^+ is not open; $L^{+\oplus} = L^{\oplus+}$.	4	8	$a \cup bb$
(9)	L is neither open nor closed; $L^\oplus$ is not closed and L^+ is not open; $L^{+\oplus} \neq L^{\oplus+}$.	5	10	$a \cup ab \cup bb$

BGS 2011: Classification of formal languages by the structure of algebra $(B(L),^+,^\oplus)$.

Questions

We ask the following questions for each of the 9 possible algebras:

(1) Can this algebra be generated by a prefix-free language?
(2) Can this algebra be generated by a regular prefix-free language?
(3) Can this algebra be generated by a regular prefix-free language of an arbitrary state complexity?
(4) What are the maximal state complexities of languages generated in this algebra by a prefix-free regular language?
(5) Is there a prefix-free regular generator which maximizes all of these complexities at the same time?

We ask the same questions for factor- and subword-free languages.

Kuratowski algebras generated by prefix-, factor-, and subword-free languages

Case	Languages in $B(L)$	State complexities: prefix-free case	Prefix-free generator
(1)	L	1 (2)	$\emptyset$ (ε)
(2)	L, L^+	n, n	✓
(5)	$L, L^+, L^\oplus$	$n, n, 1$	$\{a^{n-2}\}$ over $\{a, b\}$
(6)	$L, L^+, L^\oplus, L^{\oplus+}$	n, n, n, n	✓
(8)	$L, L^+, L^\oplus, L^{\oplus+}$	$n, n, 2^{n-3}+2, 2^{n-3}+2$	✓
(9)	$L, L^+, L^\oplus, L^{\oplus+}, L^{+\oplus}$	$n, n, 2^{n-3}+2, 2^{n-3}+2, 2^{n-3}+2$	✓

Cases (3), (4), and (7) cannot be generated by any prefix-free language

Case	Languages in $B(L)$	State complexities: factor-free case	Subword-free gener.
(1)	L	1 (2)	$\emptyset$ (ε)
(2)	L, L^+	$3, 3$	{a} over {a,b}
(5)	$L, L^+, L^\oplus$	$n, n, 1$	$\{a^{n-2}\}$ over $\{a, b\}$
(8)	$L, L^+, L^\oplus, L^{\oplus+}$	$n, n, 3, 3$	$\{a, b^{n-2}\}$

Cases (3), (4), (6), (7), and (9) cannot be generated by any factor-free language

Case (5) is generated by $\{a^{n-2}\}$ over $\{a, b\}$

The conditions in case (6):

- L is neither closed nor open
- $L^{\oplus}$ is clopen and $L^{+\oplus} = L^{\oplus}$

Theorem

The Kuratowski algebra in case (5) is generated by the prefix-free language $\{a^{n-2}\}$ *over* $\{a, b\}$*. Moreover, the maximal complexities of the languages in* $B(L) = \{L, L^{+}, L^{\oplus}\}$ *are* $(n, n, 1)$*, respectively, and they all are met by this generator.*

Proof idea.

- $L^{\oplus}$ clopen; $L^{\oplus} \subseteq L$; L is prefix-free $\Rightarrow L^{\oplus} = \emptyset$
- $L = \{a^{n-2}\}$
    - prefix-free, neither closed nor open, and $L^{+\oplus} = L^{\oplus} = \emptyset$
    - minimal DFA for L^{+} (over $\{a, b\}$) has n states. □

Case (6) cannot be generated by any factor-free language

The conditions in case (6):

1. L is neither closed nor open
2. L^+ is open
3. $L^\oplus$ is not closed
4. $L^{\oplus +} \neq L^+$

Theorem

No factor free language satisfies the conditions in case (6).

Proof idea:

- $L \neq \emptyset$ and $L \neq \{\varepsilon\}$, otherwise L is clopen.
- If $ua \in L$, then $a \in L$, otherwise L^+ is not open.
- If L is factor-free, we must have $\emptyset \neq L \subseteq \Sigma$.
- However then $L^\oplus = L$, a contradiction with L is not open. □

Summary

- the exact complexity of the star-complement-star operation on prefix-free languages
- Kuratowski algebras generated by prefix-, factor-, and subword-free languages under positive closure and complement
    - whether or not each of 9 possible algebras can be generated by a prefix-, factor, or subword-free language
    - found maximal complexities of the generated languages
    - provided a generator that maximizes all these complexities

Open problems

- the same questions for suffix-free languages
- some other closure operators
- nondeterministic case

Thank You for Your Attention