Star-Complement-Star on Prefix-Free Languages

Matúš Palmovský, Juraj Šebej

DCFS 2015

Outline

History and motivation

Prefix-free

- used in coding theory (Huffman codes, country calling codes)
- complexity of basic operations on DFAs and NFAs
  by Han, Salomaa, Wood (2009), Jirásková, Krausová (2010)
- combined operations by Han, Salomaa, Yu (2009)
- difference and left quotient by Krausová (2012)
- cyclic shift by Jirásek and Jirásková (2013)
- right quotient and reversal by Jirásek *et al.* (2014)

History and motivation

Kuratowski

Closures in Formal Languages and Kuratowski's Theorem by Brzozowski, Grant and Shallit (2011)

Star-complement-star on regular languages by Jirásková and Shallit (2012)

Trivial bound: double exponential
Lower bound: $2^{\frac{1}{8}n\log n}$
Upper bound: $2^{3n\log n}$

Prefix

Definition

$w = uv$

- u is a prefix of w;
- u is a proper prefix of w if $v \neq \varepsilon$.

Example

$w =$ Waterloo

- ε, W, Wa, Wat, Wate, Water, Waterl, Waterlo, Waterloo are prefixes of w;

Definition

A language L is prefix-free if w in $L \Rightarrow$ no proper prefix of w is in L.

Example

- $\{Wat, Waterloo\}$ is not prefix-free.
- $L \subseteq \{a, b\}^* \Rightarrow L \cdot c$ is prefix-free.

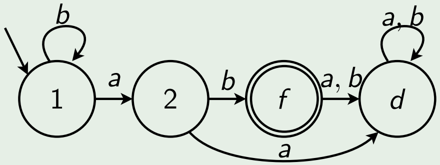
Automaton for prefix-free languages

Proposition (Han, Salomaa, Wood 2006)

Properties of minimal DFA for prefix-free languages:

- one final states f
- dead state d
- f goes to d on every symbol

Example



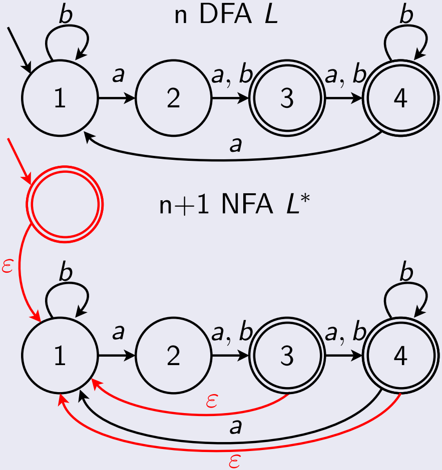
Definition

The state complexity of a regular language L ($\mathrm{sc}(L)$) is the number of states in the minimal DFA for L.

Kleene closure - L^*

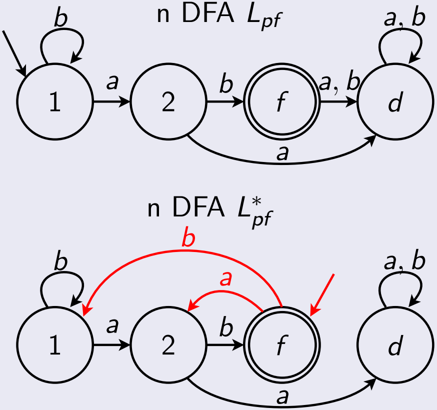
Construction of NFA for star

- ε-transitions from final states to initial state
- new initial state which is final



Construction for prefix-free

- without ε-transitions
- without new initial state



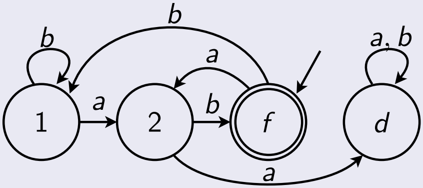
Complement - L_{pf}^{*c}

Construction of L_{pf}^{*c}

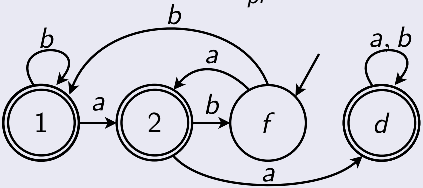
- deterministic case
- $F \leftrightarrow F^c$
- can be used on L_{pf}^{*}

$L_{pf}^{*} \rightarrow L_{pf}^{*c}$

n DFA L_{pf}^{*}



n DFA L_{pf}^{*c}

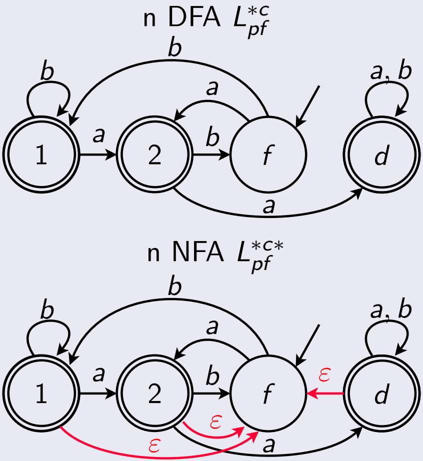


Star-complement-star - L_{pf}^{*c*}

L_{pf}^{*c*}

- L_{pf}^{*c} deterministic
- Kleene Closure of L_{pf}^{*c}
- ε-transitions from final states to initial state

$L_{pf}^{*c} \to L_{pf}^{*c*}$



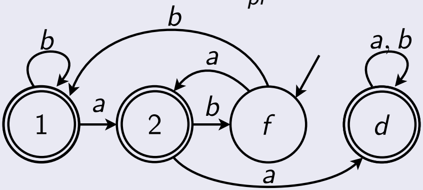
Star-complement-star - L_{pf}^{*c*}

L_{pf}^{*c*}

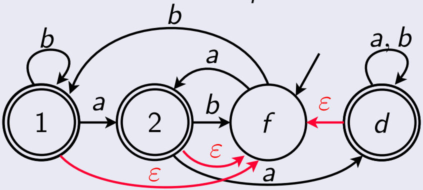
- L_{pf}^{*c} deterministic
- Kleene Closure of L_{pf}^{*c}
- ε-transitions from final states to initial state
- add new initial state which is final

$L_{pf}^{*c} \rightarrow L_{pf}^{*c*}$

n DFA L_{pf}^{*c}



n NFA L_{pf}^{*c*}



Star-complement-star - L_{pf}^{*c*}

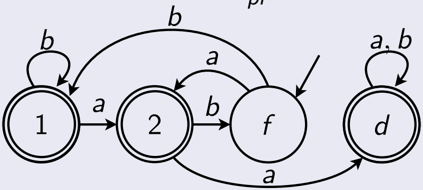
L_{pf}^{*c*}

- L_{pf}^{*c} deterministic

- Kleene Closure of L_{pf}^{*c}
- ε-transitions from final states to initial state
- ~~add new initial state which is final~~
- change of initial state
- the same transitions from 1 and f

$L_{pf}^{*c} \to L_{pf}^{*c*}$

n DFA L_{pf}^{*c}



n NFA L_{pf}^{*c*}

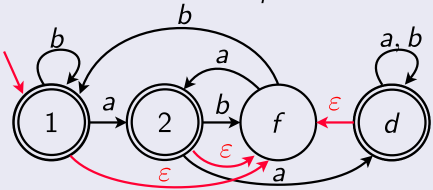


Table of frequencies of sc for L^{*c*}

List of binary, prefix-free automata.

$n/\mathrm{sc}(L^{*c*})$	1	2	3	4	5	6	7	8	9	10	AVG.
3	-	2	1	-	-	-	-	-	-	-	2.333
4	18	1	7	2	-	-	-	-	-	-	1.75
5	374	1	83	37	24	2	-	-	-	-	1.737
6	10374	1	1638	353	482	359	172	42	26	6	1.71
7	356623	1	47123	5259	7501	8194	8044	4450	2663	1867	1.738

$n/\mathrm{sc}(L^{*c*})$	11	12	13	14	15	16	17	18	AVG.
7	896	447	608	174	-	-	164	26	1.738

	max	$2^{n-3}+2$
3	3	$2^{3-3}+2=3$
4	4	$2^{4-3}+2=4$
5	6	$2^{5-3}+2=6$
6	10	$2^{6-3}+2=10$
7	18	$2^{7-3}+2=18$
8	34	$2^{8-3}+2=34$

Hypothesis

Lower bound of combined operation star-complement-star on prefix-free regular languages is $2^{n-3}+2$.

Table of frequencies of sc for L^{*c*}

List of binary, prefix-free automata.

$n/\mathrm{sc}(L^{*c*})$	1	2	3	4	5	6	7	8	9	10	AVG.
3	-	2	1	-	-	-	-	-	-	-	2.333
4	18	1	7	2	-	-	-	-	-	-	1.75
5	374	1	83	37	24	2	-	-	-	-	1.737
6	10374	1	1638	353	482	359	172	42	26	6	1.71
7	356623	1	47123	5259	7501	8194	8044	4450	2663	1867	1.738

$n/\mathrm{sc}(L^{*c*})$	11	12	13	14	15	16	17	18	AVG.
7	896	447	608	174	-	-	164	26	1.738

	max	$2^{n-3}+2$
3	3	$2^{3-3}+2=3$
4	4	$2^{4-3}+2=4$
5	6	$2^{5-3}+2=6$
6	10	$2^{6-3}+2=10$
7	18	$2^{7-3}+2=18$
8	34	$2^{8-3}+2=34$

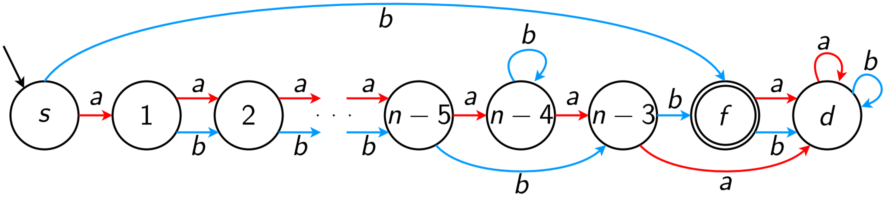
Hypothesis

Lower bound of combined operation star-complement-star on prefix-free regular languages is $2^{n-3}+2$.

Lower bound

Lemma (Lower bound)

There exists a binary regular prefix-free language L with $\mathrm{sc}(L) = n$ *such that* $\mathrm{sc}(L^{*c*}) = 2^{n-3} + 2$.



Main result: State complexity of star-complement-star

Lemma (Upper bound)

Let L be a prefix-free language over an alphabet Σ with $\mathrm{sc}(L) = n$. *Then* $\mathrm{sc}(L^{*c*}) \leqslant 2^{n-3} + 2$.

Idea of proof

- states of NFA for L_{pf}^{*c*} $\{s, 1, 2, \ldots, n-3, f, d\}$
- in corresponding subset automaton
- $d \notin S$, $f \notin S$ then $S \subseteq \{s, 1, 2, \ldots, n-3\}$ and $s \in S$
- $d \in S$, all states with d are equivalent to $\{d\}$
- $d \notin S$, $f \in S$ then S is equivalent to $S \setminus \{f\}$, except for $S = \{f\}$

Theorem (Tight upper bound)

Let L be a prefix-free language over an alphabet Σ with $\mathrm{sc}(L) = n$. *Then* $\mathrm{sc}(L^{*c*}) \leq 2^{n-3} + 2$, *and the bound is tight if* $|\Sigma| \geq 2$.

Main result: State complexity of star-complement-star

Lemma (Upper bound)

Let L be a prefix-free language over an alphabet Σ with $\mathrm{sc}(L) = n$. Then $\mathrm{sc}(L^{*c*}) \leqslant 2^{n-3} + 2$.

Idea of proof

- states of NFA for L_{pf}^{*c*} $\{s, 1, 2, \ldots, n-3, f, d\}$
- in corresponding subset automaton
- $d \notin S$, $f \notin S$ then $S \subseteq \{s, 1, 2, \ldots, n-3\}$ and $s \in S$
- $d \in S$, all states with d are equivalent to $\{d\}$
- $d \notin S$, $f \in S$ then S is equivalent to $S \setminus \{f\}$, except for $S = \{f\}$

Theorem (Tight upper bound)

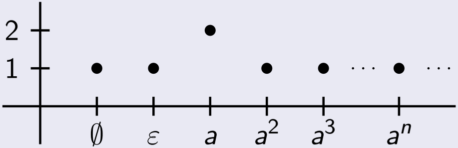
Let L be a prefix-free language over an alphabet Σ with $\mathrm{sc}(L) = n$. Then $\mathrm{sc}(L^{*c*}) \leq 2^{n-3} + 2$, and the bound is tight if $|\Sigma| \geq 2$.

Unary alphabet

Note

Every unary prefix-free language is empty set or contains one string.

Table of state complexities for L_{pf}^{*c*}



$$L_{pf}^{*c*} = \begin{cases} \varepsilon & \text{if } L = a \\ a^* & \textit{otherwise} \end{cases}$$

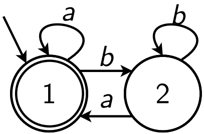
Table of complexities and special case

$n/\mathrm{sc}(L^{*c*})$	1	2	3	4	5	6	7	8	9	10	AVG.
3	-	2	1	-	-	-	-	-	-	-	2.333
4	18	1	7	2	-	-	-	-	-	-	1.75
5	374	1	83	37	24	2	-	-	-	-	1.737
6	10374	1	1638	353	482	359	172	42	26	6	1.71
7	356623	1	47123	5259	7501	8194	8044	4450	2663	1867	1.738

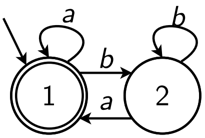
Table of complexities and special case

$n/\mathrm{sc}(L^{*c*})$	1	2	3	4	5	6	7	8	9	10	AVG.
3	-	2	1	-	-	-	-	-	-	-	2.333
4	18	1	7	2	-	-	-	-	-	-	1.75
5	374	1	83	37	24	2	-	-	-	-	1.737
6	10374	1	1638	353	482	359	172	42	26	6	1.71
7	356623	1	47123	5259	7501	8194	8044	4450	2663	1867	1.738

L^{*c*}:

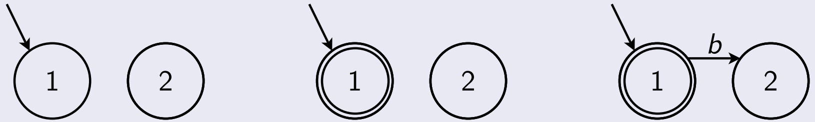


L^{*c*}:

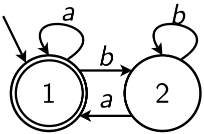


Question

How can a two state automaton for star-complement-star of prefix-free language with state complexity of at least four look?



L^{*c*}:



Theorem

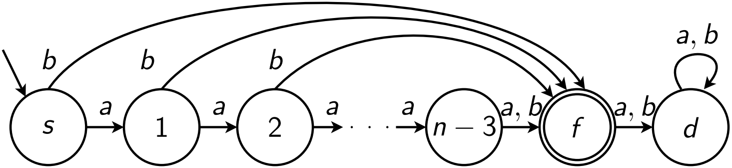
There exist exactly one binary prefix-free language with $\mathrm{sc}(L) = n$ *and* $\mathrm{sc}(L^{*c*}) = 2$.

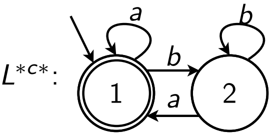
Idea of proof

- $1 \xrightarrow{b} 2 \implies b \in L$
- more then 3 states $\implies$ one of them must be reachable
- induction on $i \in \{1, n-3\}$

  assume: $\delta(i,b) = f$, $\delta(i,a) = i+1$
  - $a^{i+1}b \notin L^{*c*} \implies a^i b \in L \implies \delta(i+1,b) = f$
  - $\delta(i+1,a) = \begin{cases} i+2 & i+1 \neq n-3 \\ f & i+1 = n-3 \end{cases}$

L:





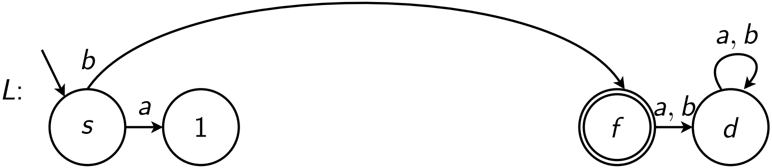
Theorem

There exist exactly one binary prefix-free language with $\mathrm{sc}(L) = n$ *and* $\mathrm{sc}(L^{*c*}) = 2$.

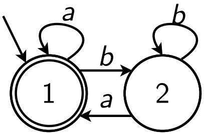
Idea of proof

- $1 \xrightarrow{b} 2 \implies b \in L$
- more then 3 states $\implies$ one of them must be reachable
- induction on $i \in \{1, n-3\}$

  assume: $\delta(i, b) = f$, $\delta(i, a) = i + 1$
  - $a^{i+1}b \notin L^{*c*} \implies a^i b \in L \implies \delta(i+1, b) = f$
  - $\delta(i+1, a) = \begin{cases} i+2 & i+1 \neq n-3 \\ f & i+1 = n-3 \end{cases}$



L^{*c*}:



Theorem

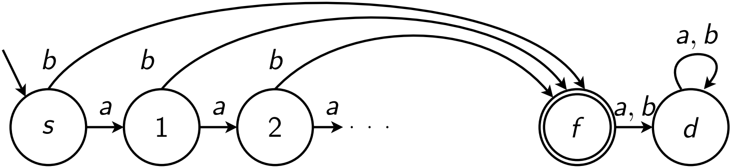
There exist exactly one binary prefix-free language with $\mathrm{sc}(L) = n$ *and* $\mathrm{sc}(L^{*c*}) = 2$.

Idea of proof

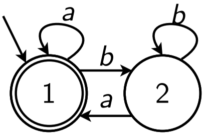
- $1 \xrightarrow{b} 2 \implies b \in L$
- more then 3 states $\implies$ one of them must be reachable
- induction on $i \in \{1, n-3\}$

  assume: $\delta(i, b) = f$, $\delta(i, a) = i + 1$
  - $a^{i+1}b \notin L^{*c*} \implies a^i b \in L \implies \delta(i+1, b) = f$
  - $\delta(i+1, a) = \begin{cases} i+2 & i+1 \neq n-3 \\ f & i+1 = n-3 \end{cases}$

L:



L^{*c*}:



Theorem

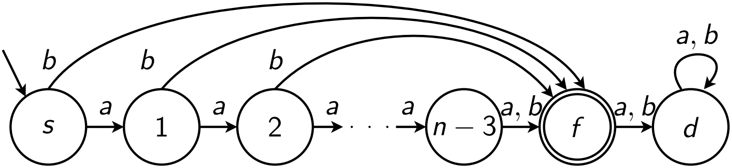
There exist exactly one binary prefix-free language with $\mathrm{sc}(L) = n$ *and* $\mathrm{sc}(L^{*c*}) = 2$.

Idea of proof

- $1 \xrightarrow{b} 2 \implies b \in L$
- more then 3 states $\implies$ one of them must be reachable
- induction on $i \in \{1, n-3\}$

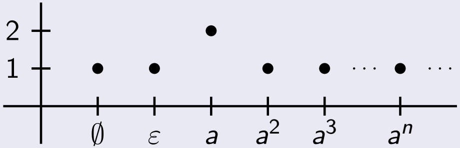
  assume: $\delta(i, b) = f$, $\delta(i, a) = i + 1$
  - $a^{i+1}b \notin L^{*c*} \implies a^i b \in L \implies \delta(i+1, b) = f$
  - $\delta(i+1, a) = \begin{cases} i+2 & i+1 \neq n-3 \\ f & i+1 = n-3 \end{cases}$

L:



Summary

- Prefix-free L with $\mathrm{sc}(L) = n \implies \mathrm{sc}(L^{*c*}) \leq 2^{n-3} + 2$
- Tightness $|\Sigma| \geq 2$
- Unary case



- There is exactly one binary prefix-free L with $\mathrm{sc}(L) = n$ such that, $\mathrm{sc}(L^{*c*}) = 2$

Open problems

- Kuratowski's 14 Theorem
- State complexity
  - Average value

n	AVG.
3	2.333
4	1.75
5	1.737
6	1.71
7	1.738
8	1.785

  - Magic numbers

$n/\mathrm{sc}(L^{*c*})$	11	12	13	14	15	16	17	18
7	896	447	608	174	-	-	164	26

$n/\mathrm{sc}(L^{*c*})$	26	27	28	29	30	31	32	33	34
8	998	-	-	8	-	-	-	1324	142

THANK YOU FOR YOUR ATTENTION