Complement on Free and Ideal Languages

Peter Mlynárčik

Slovak Academy of Science, Košice, Slovakia

Advisor: Galina Jirásková

DCFS 2015, Waterloo, Canada

Outline

1. Basic Notions and Known Facts
2. Free Languages
3. Ideal Languages
4. Open Questions

Finite Automata

Definition

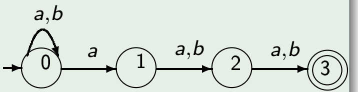
Nondeterministic finite automaton (NFA) is a five-tuple $A = (Q, \Sigma, \delta, s, F)$

- exactly one initial state s
- transition function $\delta : Q \times \Sigma \to 2^Q$

Definition

The nondeterministic state complexity of L is the number of states of minimal NFA for L. We use denotation $\mathrm{nsc}(L)$.

Example



- $\delta(0, a) = \{0, 1\}$
- $L_{3a} = \{w \in \{a, b\}^* \mid w$ has an a in the 3rd position from the end$\}$
- $\mathrm{nsc}(L_{3a}) \leq 4$

Fooling-Set Lower-Bound Method for NFAs

Definition (Fooling-Set)

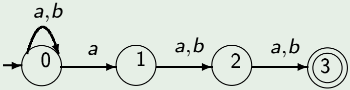
A set of pairs of strings $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$ is called a fooling set for a language L if for all i, j in $\{1, 2, \ldots, n\}$,

(F1) $x_i y_i \in L$, and

(F2) if $i \neq j$, then $x_i y_j \notin L$ or $x_j y_i \notin L$.

Example

L_{3a}:



$\{(\varepsilon, aaa), (a, aa), (aa, a), (aaa, \varepsilon)\}$ is a fooling set for L_{3a}

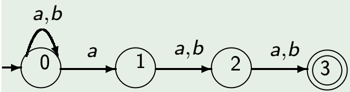
Fooling-Set Lower-Bound Method for NFAs

Lemma (Birget, 1993)

Let $\mathcal{F}$ be a fooling set for a language L.
Then every NFA for L has at least $|\mathcal{F}|$ states.

Example

$L_{3a} = \{w \in \{a, b\}^* \mid w \text{ has an } a \text{ in the 3rd position from the end } \}$



$\{(\varepsilon, aaa), (a, aa), (aa, a), (aaa, \varepsilon)\}$ is a fooling set for L_{3a}.

- a fooling set for L_{3a} with four elements $\Longrightarrow \mathrm{nsc}(L_{3a}) \geq 4$.
- there is an NFA for L_{3a} with four states $\Longrightarrow \mathrm{nsc}(L_{3a}) \leq 4$.

Hence $\mathrm{nsc}(L_{3a}) = 4$.

Finite Automata

Definition

The deterministic finite automaton (DFA) is a five-tuple $A = (Q, \Sigma, \delta, s, F)$

- transition function $\delta : Q \times \Sigma \to Q$

Definition

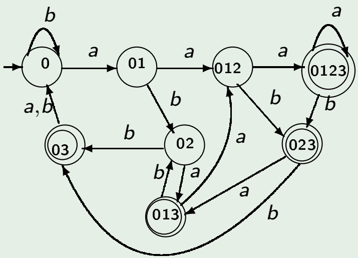
The (deterministic) state complexity of L is the number of states of minimal DFA for L. We use denotation $\mathrm{sc}(L)$.

NFA $\longrightarrow$ DFA (Rabin, Scott 1959)

Every NFA with n states has an equivalent DFA with at most 2^n states (subset construction).

Example (NFA-to-DFA)

Language L_{3a}



- a DFA constructed by subset construction
- in this case $\mathrm{sc}(L_{3a}) = 8$

Complement

Definition

Let $L \subseteq \Sigma^*$. The complement of L is $L^c = \Sigma^* \setminus L$.

DFA case - construction of DFA for complement

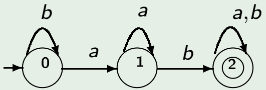
- Let A be DFA accepting a language L.
- Let DFA A^c be automaton constructed from A by interchanging final and nonfinal states.
- Then A^c accepts the complement of L.
- A is minimal $\iff$ A^c is minimal.

Complement: DFA case

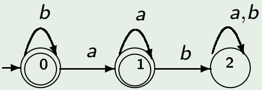
In DFA case, the number of states of minimal DFA for complement remains the same, that is,

$$\mathrm{sc}(L) = sc(L^c)$$

Example (DFA - *ab*)



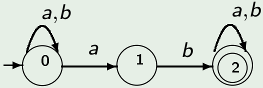
Example (DFA - no *ab*)



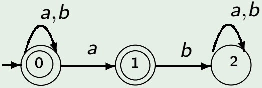
Complement: NFA case

It is not possible to get an NFA for complement from a given NFA in the same way like in DFA case:

Example (NFA - ab)



Example (NFA - $F \leftrightarrow F^c$)



NFA case - construction NFA for complement

- NFA A - accepting a language L
- DFA B - DFA constructed from A by subset construction
- DFA B^c - automaton constructed from DFA B by interchanging final and nonfinal states, it accepts L^c
- if $\mathrm{nsc}(L) = n$, then $\mathrm{nsc}(L^c) \leq 2^n$

Complement: NFA case

There are n-state NFA languages
whose complement requires 2^n nondeterministic states:

- Sakoda, Sipser (1978): $|\Sigma| = 2^n$
- Birget (1993): $|\Sigma| = 4$

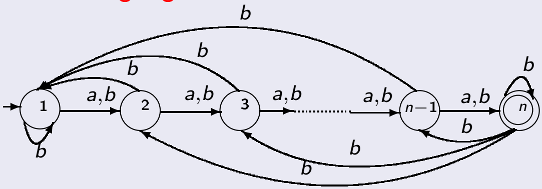
Theorem (Galina Jirásková, 2005)

Let $L \subseteq \Sigma^*$ and $\mathrm{nsc}(L) = n$.
Then $\mathrm{nsc}(L^c) \leq 2^n$, and the bound is tight if $|\Sigma| \geq 2$.

Complement: NFA case

Proof Idea.

- upper bound: for every L with $\mathrm{nsc}(L) = n$, there is an NFA for L^c with at most 2^n states
- lower bound: there is a binary L with $\mathrm{nsc}(L) = n$, such that every NFA for L^c has at least 2^n states; L - witness language



- tight upper bound: lower bound and upper bound are the same

Free Languages

Definition

$w = uxv$

- u is a prefix of w
- v is a suffix of w
- x is a factor of w

$w = u_0 v_1 u_1 v_2 u_2 \cdots v_m u_m$

- $v_1 v_2 \cdots v_m$ is a subword of w

Example

$w = \mathrm{USTAV}$

- USTA is a prefix of w
- AV is a suffix of w
- STA is a factor of w
- SAV is a subword of w

Definition

- L is prefix-free iff
  $w \in L \Rightarrow$ no proper prefix of w in L
- suffix-, factor-, subword-free defined similarly

Example

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

Motivation and History

Motivation and History

- prefix codes (Huffman coding)
- country calling codes
- Han, Salomaa (2009, 2010): suffix-free (DFA, NFA)
- Han, Salomaa, Wood (2009): prefix-free (DFA, NFA)
- Brzozowski et al. (2009,2014): ideal, closed, factor-free, subword-free (DFA)
- Jirásková, Mlynárčik (DCFS2014): prefix-free, suffix-free
  - $|\Sigma| \geq 3$: tight upper bound 2^{n-1}
  - $|\Sigma| = 2$: upper bound for prefix-free $2^{n-1} - 2^{n-3} + 1$
  - $|\Sigma| = 1$: $\mathrm{nsc}(L) = n \Longrightarrow \mathrm{nsc}(L^c) \in \Theta(\sqrt{n})$

Complement on Free Languages

Theorem (Suffix-Free Language - Binary Case)

- *upper bound:* $\mathrm{nsc}(L^c) \leq 2^{n-1} - 2^{n-3} + 2$
- *lower bound:* $2^{\lfloor \frac{n}{2} \rfloor - 1}$

(tight upper bound 2^{n-1}, if $|\Sigma| \geq 3$ (DFCS 2014))

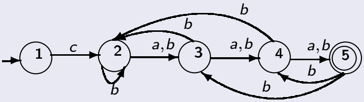
Proof Idea - Upper Bound

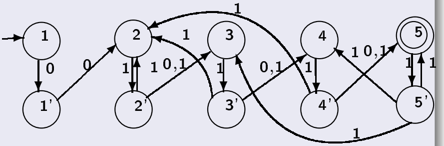
1. L - suffix-free - NFA A - n states.
2. L^R - prefix-free - NFA A^R (reverse of A) - n states.
3. $(L^R)^c$ - NFA N - at most $2^{n-1} - 2^{n-3} + 1$ states (DCFS 2014).
4. $(L^R)^c = (L^c)^R \Rightarrow$ NFA N .
5. L^c - NFA N^R (reverse of N) - at most $2^{n-1} - 2^{n-3} + 2$ states (with unique initial state).

Complement on Free Languages

Proof Idea - Lower Bound

Using homomorphism h from ternary language to binary one:
$h : c \to 00,\ a \to 10,\ b \to 11$





- ternary n-state NFA for L
- suffix-free
- $\mathcal{F}$ - fooling set for L^c, $|\mathcal{F}| = 2^{n-1}$ (DFCS 2014)

- binary $2n$-state NFA for $h(L)$
- suffix-free
- $\{(h(x), h(y)) \mid (x, y) \in \mathcal{F}\}$ - f. set for $h(L)^c$ of size 2^{n-1}

$\Longrightarrow$ lower bound: $2^{\lfloor \frac{n}{2} \rfloor - 1}$

Complement on Free Languages

Prefix-Free Language - Binary Case

- upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-1} - 2^{n-3} + 1$ (DFCS 2014)
- lower bound: $2^{\lfloor \frac{n}{2} \rfloor - 1}$

(tight upper bound 2^{n-1}, if $|\Sigma| \geq 3$ (DFCS 2014))

Factor-Free Language

- For $|\Sigma| \geq 3$, tight upper bound: $2^{n-2} + 1$
- For $|\Sigma| = 2$,
  - upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-2} - 2^{n-4} + 1$
  - lower bound: $\Omega(2^{\frac{n}{2}})$

Subword-Free Language

- upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-2} + 1$
- tight for $|\Sigma| \geq 2^{n-2}$

Complement on Free Languages-Unary

Every free unary language L can contain only one string.

$$L = \{a^n\} \Longrightarrow L^c = \{a^k \mid k \neq n\}$$

Theorem (Unary Free Language)

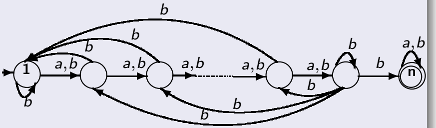
Let L be a unary prefix-free or suffix-free or factor-free or subword-free language with $\mathrm{nsc}(L) = n$. Then $\mathrm{nsc}(L^c) = \Theta(\sqrt{n})$.

Complement on Ideal Languages

Right Ideal: $L = L\Sigma^*$

upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-1}$

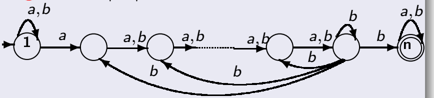
tight for $|\Sigma| \geq 2$



Two-Sided Ideal: $L = \Sigma^* L \Sigma^*$

upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-2}$

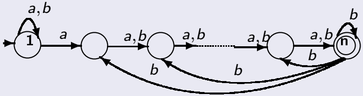
tight for $|\Sigma| \geq 2$



Left Ideal: $L = \Sigma^* L$

upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-1}$

tight for $|\Sigma| \geq 2$

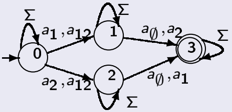


All-Sided Ideal: $L = L \sqcup\!\sqcup\, \Sigma^*$

$\sqcup\!\sqcup$ is shuffle operation

upper bound: $\mathrm{nsc}(L^c) \leq 2^{n-2}$

tight for $|\Sigma| \geq 2^{n-2}$



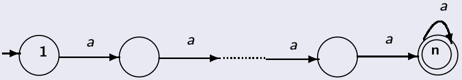
Complement on Ideal Languages

Unary Ideal

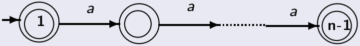
if $\mathrm{nsc}(L) = n$, then $\mathrm{nsc}(L^c) = n - 1$

$L \longrightarrow L^c$

L:



L^c:



Fooling set contains $n-1$ pairs:
$\{(\varepsilon, a^{n-2}), (a^1, a^{n-3}), \ldots, (a^i, a^{n-2-i}), \ldots, (a^{n-2}, \varepsilon)\}$

Summary - Nondeterministic Complexity of Complementation on Free Languages and Ideal Languages

CLASS	nsc	$\|\Sigma\|$	$\|\Sigma\| = 2$
suffix-free	2^{n-1}	3;not 2	$\geq 2^{\frac{n}{2}}$
prefix-free	2^{n-1}	3;not 2	$\geq 2^{\frac{n}{2}}$
factor-free	$2^{n-2}+1$	3;not 2	$\geq 2^{\frac{n}{2}}$
subword-free	$2^{n-2}+1$	2^{n-2};less?	?
unary-free	$\Theta(\sqrt{n})$		
right-ideal	2^{n-1}	2	
left-ideal	2^{n-1}	2	
two sided-ideal	2^{n-2}	2	
all sided-ideal	2^{n-2}	2^{n-2};less?	?
unary-ideal	$n-1$		

Open Questions

- possibility of improving the bounds for binary cases for **prefix-, suffix- and factor-free languages**, there is still large gap between $2^{\lfloor \frac{n}{2} \rfloor - 1}$ and $2^{n-1} - 2^{n-3} + 1$ $(2^{n-2} - 2^{n-4} + 1)$ remains still open
- complement on **subword-free and all-sided ideals:**
  smaller alphabets
  conjecture: all-sided ideals for binary alphabet - linear upper bound

THANK YOU FOR THE ATTENTION !

ĎAKUJEM ZA POZORNOSŤ !
KIITOS HUOMIOTA !
KOSZONOM A FIGYELMET !
СПАСИБО ЗА ВНИМАНИЕ !
DANKE !
MERCI !
GRAZIE !
OBRIGADO !
GAMSAHABNIDA !
ARIGATO !
DHAN'YAVADA !
...