Nondeterministic Complexity of Operations on Free and Convex Languages

Michal Hospodár Galina Jirásková Peter Mlynárčik

Slovak Academy of Science, Košice, Slovakia

Seminár Matematického ústavu SAV
Košice, 13 Júl 2017

Authors ...

Figure: Michal, Galina, Peter

Outline

Nondeterministic Finite Automata

Definition (NFA)

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

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

Nondeterministic Finite Automata

Definition (NFA)

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

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

Definition (nsc)

The nondeterministic state complexity of L is the number of states of some minimal NFA for L. We use the denotation nsc(L).

Nondeterministic Finite Automata

Definition (NFA)

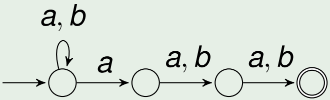
Nondeterministic finite automaton (NFA) is a quintuple $A = (Q, \Sigma, \delta, s, F)$

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

Definition (nsc)

The nondeterministic state complexity of L is the number of states of some minimal NFA for L. We use the denotation 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$\}$
- $\text{nsc}(L_{3a}) \leq 4$

Nondeterministic Finite Automata

Definition (NFA)

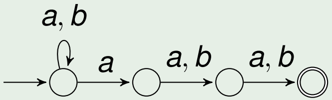
Nondeterministic finite automaton (NFA) is a quintuple $A = (Q, \Sigma, \delta, s, F)$

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

Definition (nsc)

The nondeterministic state complexity of L is the number of states of some minimal NFA for L. We use the denotation 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$\}$
- $\text{nsc}(L_{3a}) \leq 4$

If more initial states are allowed, we use the denotation NNFA

Prefix-, Suffix-, Factor-, Subword-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

$w =$ CONFERENCE

- CONFER is a prefix of w
- RENCE is a suffix of w
- FERENC is a factor of w
- CERN is a subword of w

Definition

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

Example

- $\{\varepsilon, FR, FRANCE\}$ is not prefix-free
- $\{FRANCE, PARIS\}$ is prefix-free

Properties of Free Languages

- L is prefix-free $\Rightarrow$ no out-transition from any final state
- L is suffix-free $\Rightarrow$ no in-transition to the initial state

Lemma (Sufficient conditions for an incomplete DFA to accept suffix-free language)

- *no in-transition to the initial state,*
- *single final state,*
- *no two transitions on the same symbol to any state*

Inclusions for classes of languages:
Prefix-free $\cap$ suffix-free = bifix-free
Bifix-free $\supsetneq$ factor-free $\supsetneq$ subword-free

Convex languages

Definition

- L is prefix-convex iff $u, uvw \in L \Rightarrow uv \in L$
- suffix-, factor-, subword-convex defined analogously

Every prefix-free, prefix-closed, and right ideal language is prefix-convex;
inclusions for suffix-, factor-, subword-convex languages hold analogously

Lemma (Property of Prefix-Convex Languages)

Let $D = (Q, \Sigma, \delta, s, F)$ be a DFA*. If for each final state q and each symbol a in Σ, the state $\delta(q, a)$ is final or dead, then $L(D)$ is prefix-convex.*

Why Free and Convex Languages?

Motivation and History

- Holzer, Kutrib (2003) (NFA), nsc(L) introduced
- Han, Salomaa, Wood (2009): prefix-free (DFA, NFA)
- Han, Salomaa (2010): suffix-free (DFA, NFA)
- Brzozowski et al. (2010, 2017): convex (DFA)
- P.M. (DCFS 2015): free, ideal (complement)
- M.H., G.J., P.M. (CIAA 2016): closed, ideal (NFA)

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 $x_j y_i \notin L$.

Lemma (Birget, 1992)

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

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 $x_j y_i \notin L$.

Lemma (Birget, 1992)

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

If we insist on having a single initial state, we use very useful modification of fooling-set method.

Lemma (Jirásková, Masopust, 2011)

- *$\mathcal{A}$, $\mathcal{B}$ - sets of pairs of strings*
- *u, v - two strings*
- *$\mathcal{A} \cup \mathcal{B}$, $\mathcal{A} \cup \{(\varepsilon, u)\}$, and $\mathcal{B} \cup \{(\varepsilon, v)\}$ are fooling sets for a language L.*

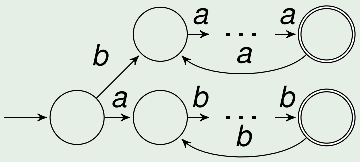
Then every NFA *with a single initial state for L has at least $|\mathcal{A}| + |\mathcal{B}| + 1$ states.*

Other Lower-Bound Methods for NFAs

Lemma (q-lemma)

Let A be an NNFA. *Let for each state q of A, the singleton set $\{q\}$ be reachable and co-reachable in A.*
Then A is minimal.

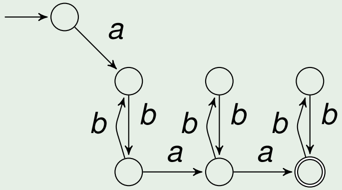
Example



Corollary

Let A be a trim NFA. *If both A and A^R are incomplete* DFAs, *then A and A^R are minimal* NFAs.

Example



We use these claims in the proofs of our results

Complexity of Operations on Free Languages

We examined the nondeterministic state complexity of:

Binary Operations

- union ($\cup$)
- intersection ($\cap$)
- concatenation ($\cdot$)

Unary Operations

- square (L^2)
- star (Kleene closure, L^*)
- reversal (L^R)
- complementation (L^c)

Known and New Results

	Prefix-free	$\|\Sigma\|$		Suffix-free	$\|\Sigma\|$	
$K \cap L$	$mn - (m + n - 2)$	2	[2]	$mn - (m + n - 2)$	2	[3]
$K \cup L$	$m + n$	2	[2]	$m + n - 1$	2	[3]
KL	$m + n - 1$	1	[2]	$m + n - 1$	1	[1]
L^2						
L^*	n	2	[2]	n	4	[1]
L^R	n	1	[2]	$n + 1$	3	[1]
L^c	2^{n-1}	3	[2]	2^{n-1}	3	[4]
		not 2	[4]		not 2	[5]

[1] Han, Salomaa DCFS 2010

[2] Jirásková, Krausová DCFS 2010

[3] Jirásková, Olejár NCMA 2009

[4] Jirásková, Mlynárčik DCFS 2014

[5] Mlynárčik DCFS 2015

Known and New Results

	Prefix-free	$\|\Sigma\|$		Suffix-free	$\|\Sigma\|$	
$K \cap L$	$mn - (m + n - 2)$	2	[2]	$mn - (m + n - 2)$	2	[3]
$K \cup L$	$m + n$	2	[2]	$m + n - 1$	2	[3]
KL	$m + n - 1$	1	[2]	$m + n - 1$	1	[1]
L^2	$2n - 1$	1		$2n - 1$	1	
L^*	n	2	[2]	n	$4 \to 2$	[1]
L^R	n	1	[2]	$n + 1$	$3 \to 2$	[1]
L^c	2^{n-1}	3	[2]	2^{n-1}	3	[4]
		not 2	[4]		not 2	[5]

[1] Han, Salomaa DCFS 2010

[2] Jirásková, Krausová DCFS 2010

[3] Jirásková, Olejár NCMA 2009

[4] Jirásková, Mlynárčik DCFS 2014

[5] Mlynárčik DCFS 2015

Our Results

	Factor-free	$\|\Sigma\|$	Subword-free	$\|\Sigma\|$
$K \cap L$	$mn - 2(m+n-3)$	2	$mn - 2(m+n-3)$	$m+n-5$
$K \cup L$	$m+n-2$	2	$m+n-2$	2
KL	$m+n-1$	1	$m+n-1$	1
L^2	$2n-1$	1	$2n-1$	1
L^*	$n-1$	1	$n-1$	1
L^R	n	1	n	1
L^c	$2^{n-2}+1$	3 not 2	$2^{n-2}+1$	2^{n-2} smaller?

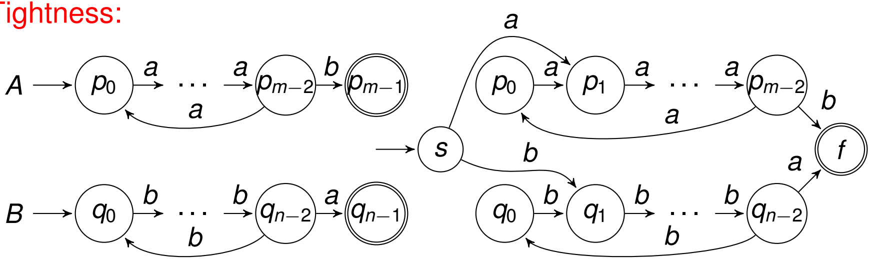
The results for complementation are from P.M., DCFS 2015

Union on Prefix-Free Languages: $m + n$

Proof Idea:

Upper Bound: merge final states and add initial state

Tightness:



Use AB-Lemma with

$\mathcal{A} = \{(a^i, a^{m-2-i}b) \mid 1 \leq i \leq m-2\} \cup \{(a^{m-1}, a^{m-2}b), (a^{m-2}b, \varepsilon)\}$

$\mathcal{B} = \{(b^j, b^{n-2-j}a) \mid 1 \leq i \leq n-2\} \cup \{(b^{n-1}, b^{n-2}a)\}$

$u = b^{n-2}a$

$v = a^{m-2}b$

Our Results on Convex Languages

	Prefix-convex	$\lvert\Sigma\rvert$	Suffix-convex	$\lvert\Sigma\rvert$	Factor-convex	$\lvert\Sigma\rvert$	Subword-convex	$\lvert\Sigma\rvert$
$K \cap L$	mn	2	·	2	·	2	·	2
$K \cup L$	$m+n+1$	2	·	2	·	2	·	2
KL	$m+n$	3	·	3	·	3	·	3
L^2	$2n$	3	·	3	·	3	·	3
L^*	$n+1$	2	·	2	·	2	·	2
L^R	$n+1$	2	·	3	·	3	·	$2n-2$

- nsc of operations on convex languages
- all upper bounds are met by ideal languages (star) or closed languages (all the other operations)

Our Results on Convex Languages

	Prefix-convex	$\|\Sigma\|$	Suffix-convex	$\|\Sigma\|$	Factor-convex	$\|\Sigma\|$	Subword-convex	$\|\Sigma\|$
$K \cap L$	mn	2	·	2	·	2	·	2
$K \cup L$	$m+n+1$	2	·	2	·	2	·	2
KL	$m+n$	3	·	3	·	3	·	3
L^2	$2n$	3	·	3	·	3	·	3
L^*	$n+1$	2	·	2	·	2	·	2
L^R	$n+1$	2	·	3	·	3	·	$2n-2$
L^c	2^n	2	$\geq 2^{n-1}+1$ $\leq 2^n$	2	· ·	2	· ·	2^n

- nsc of operations on convex languages
- all upper bounds are met by ideal languages (star) or closed languages (all the other operations) , except for complementation

Complementation on Suffix-Convex Languages

Tight upper bounds:

- suffix-closed: $2^{n-1} + 1$
- left ideal, suffix-free: 2^{n-1}

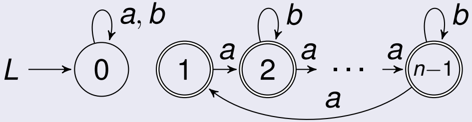
Lemma (Complementation)

If all subsets are reachable and co-reachable, then $\mathrm{nsc}(L^c) = 2^n$

Lemma (Property of Prefix-Convex Languages)

Let $D = (Q, \Sigma, \delta, s, F)$ *be a* DFA. *If for each final state* q *and each symbol* a *in* Σ*, the state* $\delta(q, a)$ *is final or dead, then* $L(D)$ *is prefix-convex.*

Suffix-convex witness meeting 2^n



$0 \cdot c = \{0, 1, \ldots, n-1\}$,
$0 \cdot d = \{1, 2, \ldots, n-1\}$,
$q \cdot e = \{n-1\}$ for each state q of A

Proof Idea

- show that L^R is prefix-convex by Lemma (Property)
- use Lemma (Complementation)

Unary Case

- Unary free languages: $L = \{a^{n-1}\} \Leftrightarrow \text{nsc}(L) = n$
- Unary convex languages:
  - $L = \{a^i \mid i \geq k\} \Rightarrow \text{nsc}(L) = k + 1$
  - $L = \{a^i \mid k \leq i \leq \ell\} \Rightarrow \text{nsc}(L) = \ell + 1$

Unary	$K \cap L$	$K \cup L$	KL	L^2	L^*	L^c
free	n; $m = n$	$\max\{m, n\}$	$m + n - 1$	$2n - 1$	$n - 1$	$\Theta(\sqrt{n})$
convex	$\max\{m, n\}$	$\max\{m, n\}$	$m + n - 1$	$2n - 1$	$n - 1$	$n + 1$
regular	mn; $(m, n) = 1$	$m + n + 1$; $(m, n) = 1$	$\geq m + n - 1$ $\leq m + n$	$\geq 2n - 1$ $\leq 2n$	$n + 1$	$2^{\Theta(\sqrt{n \log n})}$

Summary and Open Problems

The most important results:

- intersection on subword-free languages:
  $mn - 2(m + n - 3)$, $|\Sigma| = m + n - 5$
- union on prefix-free languages:
  $m + n$, $|\Sigma| = 2$
- complementation on suffix-convex languages:
  2^n, $|\Sigma| = 5$

Tight upper bounds were provided for all other combinations of operations and classes except for complementation on factor-convex and subword-convex languages (open problem)

Possible to decrease alphabet size

- subword-free
  - intersection (binary, $m = n$)
  - complementation
- subword-convex
  - reversal

Thank You for Attention

Merci beaucoup pour votre attention