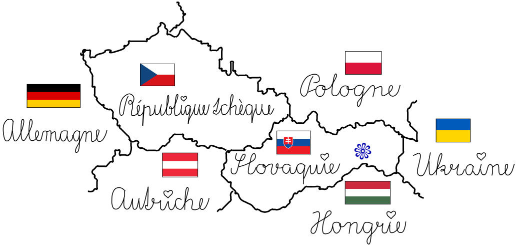
Operations on Unambiguous Finite Automata

Galina Jirásková

Mathematical Institute, Slovak Academy of Sciences, Košice, Slovakia



Joint work with Jozef Jirásek, Jr., and Juraj Šebej

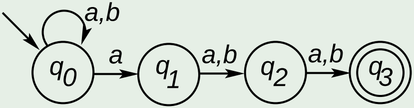
DLT 2016, Montréal, Québec, Canada

Nondeterministic and Deterministic Finite Automata

NFA $N = (Q, \Sigma, \delta, I, F)$:

- $\delta \subseteq Q \times \Sigma \times Q$
- computation on $w = a_1 a_2 \cdots a_k$
  $q_0 \xrightarrow{a_1} q_1 \xrightarrow{a_2} q_2 \xrightarrow{a_3} \cdots \xrightarrow{a_k} q_k$
  $q_0 \in I$
- accepting if $q_k \in F$
- rejecting if $q_k \notin F$

Example (An NFA)



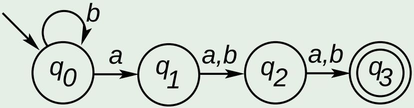
$w = aaa$

- $q_0 \xrightarrow{a} q_1 \xrightarrow{a} q_2 \xrightarrow{a} q_3$ (acc.)
- $q_0 \xrightarrow{a} q_0 \xrightarrow{a} q_0 \xrightarrow{a} q_0$ (rej.)

NFA $N = (Q, \Sigma, \delta, I, F)$ is a DFA:

- $|I| = 1$
- if (q, a, p) and (q, a, r) are in δ, then $p = r$

Example (An incomplete DFA)



- NFAs may have multiple initial states
- DFAs may be incomplete

Subset Automaton and Reverse of NFA

Definition

The (incomplete) subset automaton of NFA $N = (Q, \Sigma, \delta, I, F)$ is the DFA $(2^Q \setminus \{\emptyset\}, \Sigma, \delta', I, F') \ldots$

Proposition

Every n-state NFA can be simulated by an $(2^n - 1)$-state incomplete DFA.

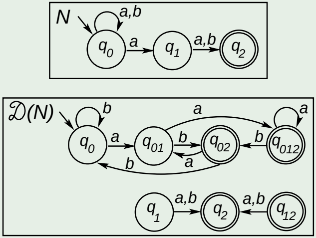
Definition

The reverse of an NFA $N = (Q, \Sigma, \delta, I, F)$ is the NFA

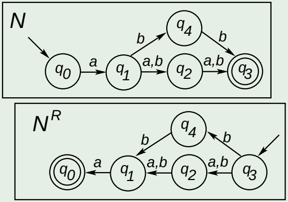
$$N^R = (Q, \Sigma, \delta^R, F, I),$$

where $(p, a, q) \in \delta^R$ iff $(q, a, p) \in \delta$

Example (Subset automaton)



Example (Reverse of NFA)



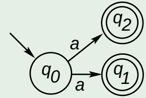
Unambiguous Finite Automata

Definition ($N = (Q, \Sigma, \delta, I, F)$)

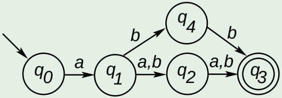
An NFA is unambiguous if it has at most one accepting computation on every input string.

- $S \subseteq Q$ is reachable in N if $S = \delta(I, w)$ for some w
- $S \subseteq Q$ is co-reachable in N if S is reachable in N^R

Example (not unambiguous)



– two accepting computations on a



– two accepting computations on abb

Proposition

An NFA is unambiguous iff

$$|S \cap T| \leq 1$$

for each reachable S and each co-reachable T

Example (unambiguous)

- (in)complete DFA
- NFA N s.t. N^R deterministic
- NFA in the first slide

Why Unambiguous Finite Automata?

Motivation and History

- fundamental notion in the theory of variable-length codes [Bersten, Perrin, Reutenauer: Codes and Automata]
- ambiguity in CF languages: ambiguous, unambiguous, and deterministic CF languages are all different
- ambiguity in finite automata [Schmidt 1978]
  - lower bound method based on ranks of matrices
- elaborated in [Leung 2005]
  - UFA-to-DFA conversion: 2^n
  - NFA-to-UFA conversion: $2^n - 1$
- lower bound method further elaborated in 2002 by Hromkovič, Seibert, Karhumäki, Klauck & Schnitger

Why Operations on Unambiguous Finite Automata?

Motivation for me:-)

- conference trip at DLT 2008 (Kyoto): A. Okhotin - ...
  "What is the complexity of complementation on UFAs?"
- operations on unary UFAs investigated by him in 2012
  - lower bound $n^{2-o(1)}$ for complementation
- the second problem for which
  "give me a large enough alphabet" method didn't work ...

Lower Bounds Methods I

Well known: To prove that a DFA is minimal, show that

- all its states are reachable, and
- no two distinct states are equivalent.

Well known(?): To prove that an NFA is minimal, describe

a fooling set for the accepted language.

For UFAs: rank of matrices [Schmidt 78, Leung 05]:

Let N be an NFA. Let M_N be the matrix in which

- rows indexed by non-empty reachable sets
- columns indexed by non-empty co-reachable sets
- in entry (S, T) we have $0/1$ if S and T are/are not disjoint.

Then every UFA for $L(N)$ has at least $\text{rank}(M_N)$ states.

Lower Bounds Methods II

Lemma (Leung 1998, Lemma 3)

Let M_n be the $(2^n - 1) \times (2^n - 1)$ matrix with

- *rows and columns indexed by non-empty subsets of $\{1, 2, \ldots, n\}$*
- *$M_n(S, T) = 0/1$ iff S and T are/are not disjoint.*

Then $rank(M_n) = 2^n - 1$.

Corollary

If each non-empty set is co-reachable in NFA N,
then every UFA equivalent to N
has $\geq$ |non-empty reachable| states.

Maslov 1970

Dokl. Akad. Nauk SSSR
Tom 194 (1970), No. 6

Soviet Math. Dokl.
Vol. 11 (1970), No. 5

ESTIMATES OF THE NUMBER OF STATES OF FINITE AUTOMATA

519.95

A. N. MASLOV

It is well known that, if $T(A)$ and $T(B)$ are representable in automata A and B with m and n states, respectively $(m \geq 1, n \geq 1)$, then:

1) $T(A) \cup T(B)$ is representable in an automaton with $m \cdot n$ states;

2) $T(A) \cdot T(B)$ is representable in an automaton with $(m-1) \cdot 2^n + 2^{n-1}$ states $(n \geq 3)$;

3) $T(A)^*$ is representable in an automaton with $(3/4) \cdot 2^m - 1$ states $(m \geq 2)$.

Let us construct examples of automata over the alphabet $\Sigma = \{0, 1\}$ for which these estimates are attained.

1. Union. A has states $\{S_0, \cdots, S_{m-1}\}$ and transitions $S_{m-1}1 = S_0$, $S_i1 = S_{i+1}$ for $i \neq m-1$, $S_i0 = S_i$, and S_{m-1} is the terminal state. B has states $\{P_0, \cdots, P_{n-1}\}$ and transitions $P_i1 = P_i$, $P_{n-1}0 = P_0$; $P_i0 = P_{i+1}$ for $i \neq n-1$, P_{n-1} is the terminal state.

2. Product. B has the states $\{P_0, \cdots, P_{n-1}\}$ and transitions $P_{n-1}1 = P_{n-2}$, $P_{n-2}1 = P_{n-1}$, $P_i1 = P_i$ for $i < n-2$, $P_{n-1}0 = P_{n-1}$, $P_i0 = P_{i+1}$ for $i \neq n-1$. P_{n-1} is the terminal state. The automaton A is the same as in the case of the union.

3. Iteration. A has the states $\{S_0, \cdots, S_{m-1}\}$ and transitions $S_{m-1}1 = S_0$, $S_i1 = S_{i+1}$ for $i \neq m-1$, $S_00 = S_0$, $S_i0 = S_{i-1}$ for $i > 0$. S_{m-1} is the terminal state.

Corresponding to A and B we construct automata as in [2,4] and we find the required number of attainable and distinct states, which proves the minimality [3].

A General Formulation of the Problem

Maslov 1970

A general formulation of the problem is as follows: We have events $T(A_i)$ $(1 \leq i \leq k)$ representable in automata A_i with n_i states, respectively, and a k-place operation f on events, preserving representability in finite automata. What is the maximal number of states of a minimal automaton representing $f(T(A_1), \ldots, T(A_k))$, for the given n_i?

"We have languages $L(A_i)$ $(1 \leq i \leq k)$
recognized by automata A_i with n_i states, respectively,
and a k-ary regular operation f.
What is the maximal number of states
of a minimal automaton recognizing $f(L(A_1), \ldots, L(A_k))$,
for the given n_i?"

In this paper:

- automata are unambiguous (UFAs)
- f: intersection, reversal, shuffle, star and positive closure,
  left and right quotients, concatenation,
  complementation, and union

Intersection on Unambiguous Finite Automata

Intersection:
$K \cap L = \{w \mid w \in K \text{ and } w \in L\}$

Known results for intersection:

DFA:	mn	binary	[Maslov 1970]
NFA:	mn	binary	[Holzer & Kutrib 2003]

Our result for intersection on UFAs:

UFA: mn $|\Sigma| \geq 2$

Proof sketch:

- upper bound: given UFAs A and B, construct the direct product automaton $A \times B$; it is a UFA
- lower bound: the witnesses in [HK'03] for NFA intersection are deterministic, so UFAs □

Shuffle on Unambiguous Finite Automata

Shuffle:

$K \sqcup\!\sqcup L = \{u_1 v_1 u_2 v_2 \cdots u_k v_k \mid u_1 u_2 \cdots u_k \in K \text{ and } v_1 v_2 \cdots v_k \in L\}$

Known results for shuffle:

DFA:	???		
in-DFA:	$2^{mn} - 1$	5-letter	[Câmpeanu, Salomaa & Yu 2002]
NFA:	mn	binary	[G. J. & Masopust, DLT 2010]

Our result for shuffle on UFAs:

UFA: $2^{mn} - 1$ $|\Sigma| \geq 5$

Proof sketch for lower bound:

- take the witness incomplete DFAs from [CSY'02]
- in the mn-state NFA for shuffle
  - each non-empty set is reachable [CSY'02]
  - each non-empty set is co-reachable □

Concatenation on Unambiguous Finite Automata

Concatenation:
$KL = \{uv \mid u \in K \text{ and } v \in L\}$

Known results for concatenation:

DFA:	$(m - 1/2) \cdot 2^n$	binary	[Maslov 1970]
NFA:	$m + n$	binary	[Holzer & Kutrib 2003]

Our result for concatenation on UFAs:

UFA: $(3/4) \cdot 2^m \cdot 2^n - 1$ $\quad |\Sigma| \geq 7$

Proof idea for the upper bound:

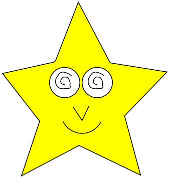
- construct an $(m + n)$-state NFA N for KL
- show that at most $(3/4) \cdot 2^m \cdot 2^n - 1$ subsets are reachable in the subset automaton of N $\square$

Star on Unambiguous Finite Automata

Star:
$L^* = \{u_1 u_2 \cdots u_k \mid k \geq 0 \text{ and } u_i \in L \text{ for all } i\}$

Known results for the star operation:

DFA:	$(3/4) \cdot 2^n$	binary	[Yu, Zhuang & K. Salomaa 1994]
NFA:	$n + 1$	unary	[Holzer & Kutrib 2003]



Our result for star on unambiguous automata:

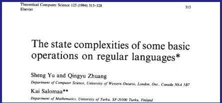
UFA: $(3/4) \cdot 2^n$ $|\Sigma| \geq 3$

Proof idea for the lower bound:

- start with YZS'94 binary witness DFA for star
- define a new symbol c
- compute the rank of the corresponding matrix □

Ternary Witness UFA for Star Meeting the Bound $(3/4) \cdot 2^n$

Yu, Zhuang & K. Salomaa 1994


Theoretical Computer Science 125 (1994) 315–328
Elsevier
315

The state complexities of some basic operations on regular languages*

Sheng Yu and Qingyu Zhuang
Department of Computer Science, University of Western Ontario, London, Ont., Canada N6A 5B7
Kai Salomaa**
Department of Mathematics, University of Turku, SF-20500 Turku, Finland

Theorem 3.3. *For any integer $n \geqslant 2$, there exists a DFA A of n states such that any DFA accepting $(L(A))^*$ needs at least $2^{n-1}+2^{n-2}$ states.*

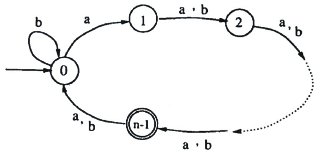
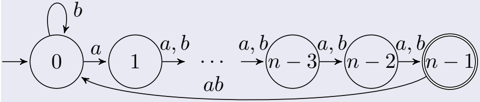
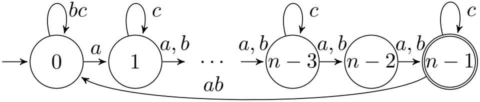


Fig. 4. DFA A_n.



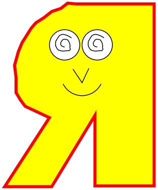


Reversal on Unambiguous Finite Automata

Reversal:
$L^R = \{w^R \mid w \in L\}$, where w^R is the mirror image of w

Known results for the reversal operation:

DFA:	2^n	binary	[Leiss 1981, Šebej 2009]
NFA:	$n+1$	binary	[Holzer & Kutrib 2003, G. J. 2005]



Reversal on UFAs:

UFA: n $|\Sigma| \geq 1$

Proof.

If A is unambiguous, then A^R is unambiguous. □

Complementation on UFAs: Partial Results

Known results for complementation:

DFA:	n	unary	[folklore]
NFA:	2^n	binary	[Birget 1993, G. J. 2005]
UFA:	$\geq n^{2-o(1)}$	unary	[Okhotin 2012]

Our unsuccessful attempts for UFAs:

- the matrix method didn't work:
  $\text{rank}(M_{L^c}) = \text{rank}(M_L) \pm 1$
- the fooling-set method didn't work:
  - if L is accepted by an n-state UFA, then every fooling set for L^c is of size $\leq n^2/2$
  - we only found a fooling set of size $n + \sqrt{n}$
  - conjecture: every fooling set for L^c is of size $\leq 2n$
- large alphabets didn't work either

Complementation on UFAs: Partial Results

Known results for complementation:

DFA:	n	unary	[folklore]
NFA:	2^n	binary	[Birget 1993, G. J. 2005]
UFA:	$\geq n^{2-o(1)}$	unary	[Okhotin 2012]

Our upper bound on complementation for UFAs:

UFA: $\leq 2^{0.79n+\log n}$

Proof sketch for the upper bound:

If L is accepted by an n-state UFA A, then

- $\text{usc}(L^c) \leq |\mathcal{R}|$ (reachable in A)
- $\text{usc}(L^c) \leq |\mathcal{C}|$ (co-reachable in A)
- if $\max\{|S| \mid S \in \mathcal{R}\} \geq n/2$, then $|\mathcal{C}|$ is small
- otherwise, $\min\{|\mathcal{R}|, |\mathcal{C}|\}$ is small $\square$

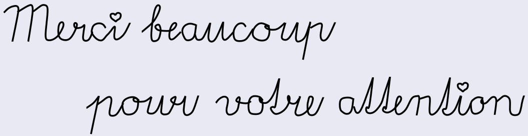
Summary and Open Problems

The complexity of operations on unambiguous finite automata:

	sc	$\|\Sigma\|$	usc	$\|\Sigma\|$	nsc	$\|\Sigma\|$
intersection	mn	2	mn	2	mn	2
left quotient	$2^n - 1$	2	$2^n - 1$	2	$n+1$	2
positive closure	$\frac{3}{4} \cdot 2^n - 1$	2	$\frac{3}{4} \cdot 2^n - 1$	3	n	1
star	$\frac{3}{4} \cdot 2^n$	2	$\frac{3}{4} \cdot 2^n$	3	$n+1$	1
shuffle	?		$2^{mn} - 1$	5	mn	2
reversal	2^n	2	n	1	$n+1$	2
concatenation	$(m - 1/2) \cdot 2^n$	2	$\frac{3}{4} \cdot 2^{m+n} - 1$	7	$m+n$	2
right quotient	n	1	$2^n - 1$	2	n	1
complementation	n	1	$\leq 2^{0.79n + \log n}$		2^n	2
			$\geq n^{2-o(1)}$	1		

Acknowledgments

1. Thank you very much for your attention



2. Many thanks to ...

- "big" Jozko and "small" Jozko
- Maria, Jonas, and Dominik
- ...

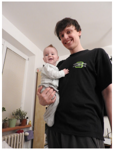
Greetings from Maria, Jonas, and Dominik

Maria 2004

Sept. 2015

3 weeks

3 months



6 months

Easter 2016

last week

Summary and Open Problems

The complexity of operations on unambiguous finite automata:

	sc	$\lvert\Sigma\rvert$	usc	$\lvert\Sigma\rvert$	nsc	$\lvert\Sigma\rvert$
intersection	mn	2	mn	2	mn	2
left quotient	$2^n - 1$	2	$2^n - 1$	2	$n + 1$	2
positive closure	$\frac{3}{4} \cdot 2^n - 1$	2	$\frac{3}{4} \cdot 2^n - 1$	3	n	1
star	$\frac{3}{4} \cdot 2^n$	2	$\frac{3}{4} \cdot 2^n$	3	$n + 1$	1
shuffle	?		$2^{mn} - 1$	5	mn	2
reversal	2^n	2	n	1	$n + 1$	2
concatenation	$(m - 1/2) \cdot 2^n$	2	$\frac{3}{4} \cdot 2^{m+n} - 1$	7	$m + n$	2
right quotient	n	1	$2^n - 1$	2	n	1
complementation	n	1	$\leq 2^{0.79n + \log n}$		2^n	2
			$\geq n^{2 - o(1)}$	1		