Syntactic NFA Minimization via SAT solving

Bastian Kauschke

November 29, 2021

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

GOAL

minimal nfas

Given an nfa M with states Q, nfa N with states S is a *union nfa* of M if

- each $s \in S$ accepts $\bigcup\{L(N, q) : q \in X\}$ with $X \subseteq Q$
- $L(N) = L(M)$

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Given an nfa M with states Q, nfa N with states S is a *union nfa* of M if

- each $s \in S$ accepts $\bigcup\{L(N, q) : q \in X\}$ with $X \subseteq Q$
- $L(N) = L(M)$

An nfa N is *rpd* if N^R is a partial dfa.
The states of rpd nfas accept pairwise disjoint languages.
All trimmed nfas whose states accept pairwise disjoint languages are rpd.

nfa $N = (Q, \delta, I, F)$ accepting L is **atomic** if the following equivalent statements hold

- $\mathrm{rsc}(N^R)$ is a minimal dfa
- each $q \in Q$ accepts a language from $\mathrm{BLD}(L)$, the closure of $\mathrm{LD}(L)$ under all set-theoretic boolean operations
- each $q \in Q$ accepts a union of congruence classes of the *Nerode left congruence* $\sim_L$:

$$\begin{aligned} u \sim_L v &\iff \forall x \in \Sigma^*. u \in x^{-1}L = v \in x^{-1}L \\ &\iff (u^r)^{-1}L^r = (v^r)^{-1}L^r \end{aligned}$$

- N is a union nfa of the átomaton $\mathbf{dfa}(L^r)^R$

nfa $N = (Q, \delta, I, F)$ accepting L is **subatomic** if the following equivalent statements hold

- the transition monoid of $\mathrm{rsc}(N^R)$ is isomorphic to $\mathrm{syn}(L^r)$
- each $q \in Q$ accepts a language from BLRD(L), the closure of LD(L) under all boolean operations and right derivatives
- each $q \in Q$ accepts a union of congruence classes of the *syntactic congruence* $\equiv_L$:

$$u \equiv_L v \iff \forall x, y \in \Sigma^*. u \in x^{-1} L y^{-1} = v \in x^{-1} L y^{-1}$$

- N is a union nfa of the syntactic nfa **syn**(L)

given **dfa**(L) and an rpd nfa $M = (Q, \delta, I, \{q_f\})$ accepting L,

does a union nfa N of M with k states exist?

We are searching for an nfa N with states $\mathbf{k}$:

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG

We are searching for an nfa N with states $\mathbf{k}$:

- N is a union nfa: $H \subseteq \mathbf{k} \times Q$

$$H[i] = X \iff L(N, i) = \bigcup \{L(M, q) : q \in X\}$$

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

We are searching for an nfa N with states $\mathbf{k}$:

- N is a union nfa: $H \subseteq \mathbf{k} \times Q$

$$H[i] = X \iff L(N, i) = \bigcup \{L(M, q) : q \in X\}$$

- N accepts L: $G \subseteq \mathrm{LD}(L) \times \mathbf{k}$

$$G[u^{-1}L] = Y \iff u^{-1}L = \bigcup \{L(N, i) : i \in Y\}$$

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

We are searching for an nfa N with states $\mathbf{k}$:

- N is a union nfa: $H \subseteq \mathbf{k} \times Q$

$$H[i] = X \iff L(N,i) = \bigcup\{L(M,q) : q \in X\}$$

- N accepts L: $G \subseteq \mathrm{LD}(L) \times \mathbf{k}$

$$G[u^{-1}L] = Y \iff u^{-1}L = \bigcup\{L(N,i) : i \in Y\}$$

As a commutative diagram:

$$\begin{tikzcd}
\mathrm{LD}(L) \arrow[d, "G"'] \arrow[dr, "\supseteq"] & \\
\mathbf{k} \arrow[r, "H"'] & Q
\end{tikzcd}$$

Require states to accept their assigned language using the transition relations $R_a \subseteq \mathbf{k} \times \mathbf{k}$ $(a \in \Sigma)$:

$$\begin{tikzcd} \mathrm{LD}(L) \arrow[d, "G"'] \arrow[dr, "\supseteq"] & \\ \mathbf{k} \arrow[r, "H"'] & Q \end{tikzcd} \qquad \begin{tikzcd} Q \arrow[r, "\delta_a"] & Q \\ \mathbf{k} \arrow[u, "H"] \arrow[r, "R_a"'] & \mathbf{k} \arrow[u, "H"'] \end{tikzcd}$$

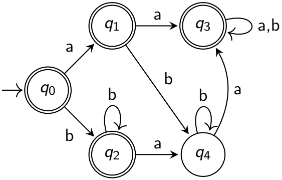
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

Require states to accept their assigned language using the transition relations $R_a \subseteq \mathbf{k} \times \mathbf{k}$ ($a \in \Sigma$):

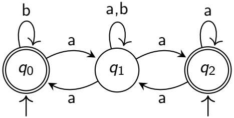
$$\begin{tikzcd}
\mathrm{LD}(L) \arrow[d, "G"'] \arrow[dr, "\supseteq"] & \\
\mathbf{k} \arrow[r, "H"'] & Q
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
Q \arrow[r, "\delta_a"] & Q \\
\mathbf{k} \arrow[u, "H"] \arrow[r, "R_a"'] & \mathbf{k} \arrow[u, "H"']
\end{tikzcd}$$

$$N = (\mathbf{k}, (R_a)_{a\in\Sigma}, G[L], H^{-1}[q_f])$$

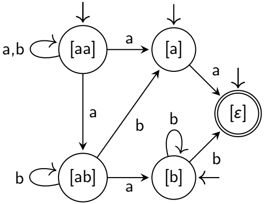
$$L = \{w \in \Sigma^* : |w|_b = 0 \vee |w|_a \neq 1\}$$



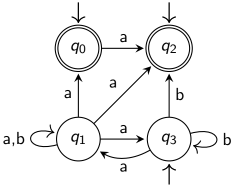
dfa(L)



N

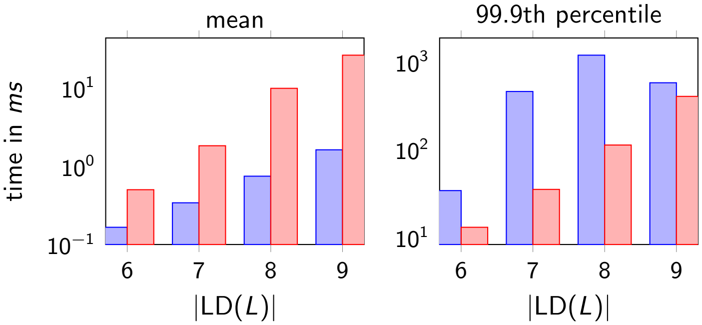


syn(L)

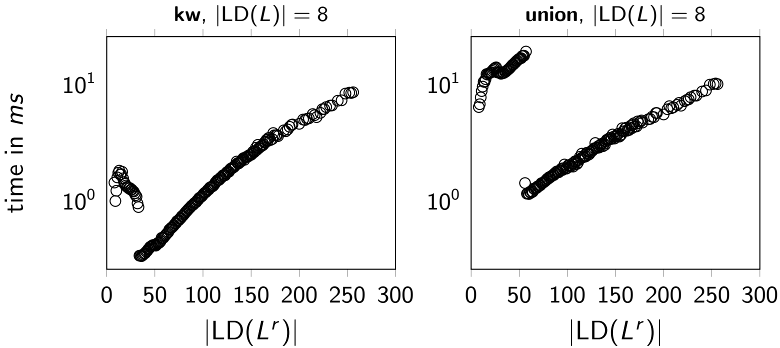


N_{syn}

performance results



performance results



size results

k	total	$\mathbf{ns}(L) < \lvert\mathrm{LD}(L)\rvert$	$\mathbf{ns}(L) < \mathbf{natm}(L)$	$\mathbf{ns}(L) < \mathbf{nsyn}(L)$	unknown
1	**2**	**1** (50.0%)	**0** (0.0%)	**0** (0.0%)	0
2	**24**	**3** (12.5%)	**0** (0.0%)	**0** (0.0%)	0
3	**1028**	**123** (12.0%)	**0** (0.0%)	**0** (0.0%)	0
4	**56014**	**5911** (10.6%)	**86** (0.15%)	**86** (0.15%)	0
5	**3705306**	**335820** (9.1%)	**13376** (0.36%)	**11122** (0.30%)	0
6	1269000	98376 (7.8%)	5399 (0.43%)	4540 (0.36%)	0
7	1000000	67904 (6.8%)	4159 (0.42%)		93
8	1000000	60497 (6.0%)	4084 (0.41%)		362
9	1000000	55131 (5.5%)	3668 (0.37%)		372
10	1000000	51563 (5.2%)	3598 (0.36%)		318
11	1000000	48070 (4.8%)	3465 (0.34%)		241