The State Complexity of Distinguishability Operation Combined with the Union Operation

Cezar Câmpeanu*

18.08.2017

*School of Mathematical and Computational Sciences
University of Prince Edward Island, CANADA

Minimizing DFAs

Checking Minimality

For a DFA to check if it is minimal we can:

- Try to minimize ($O(n \log n)$).
- Find a pair of equivalent states ($O(n^2)$)
- In case we know the language we can test if some words distinguish states in the DFA – although expensive, humans often use this approach to check minimality.

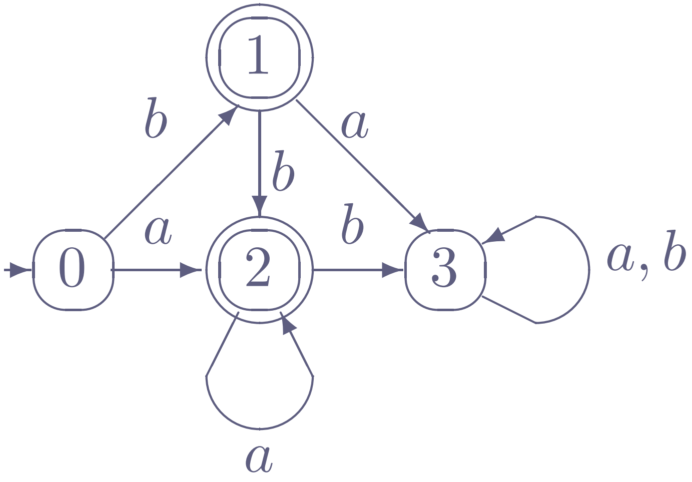

Figure 1: ε distinguishes 1 and 2 from 0 and 3 and a distinguishes 1 from 2 and 0 from 3.

Distinguishability Language

- For a regular language L:

$$\mathsf{D}(L) = \{w \mid \exists x, y \in \Sigma^\star \ (xw \in L \text{ and } yw \notin L)\}.$$

- For words

$$\mathsf{D}_L(x,y) = \{w \mid xw \in L \not\Leftrightarrow yw \in L\}\,.$$

- For $\mathcal{A} = \langle Q, \Sigma, \delta, q_0, F \rangle$,

$$\mathsf{D}(\mathcal{A}) = \{w \mid \exists p, q \in Q \ (\delta(p,w) \in F \text{ and } \delta(q,w) \notin F)\}.$$

- For states

$$\mathsf{D}_L(p,q) = \{w \mid (\delta(p,w) \in F \text{ and } \delta(q,w) \notin F)\}.$$

Properties of D

Basic Properties

- If $A_1 \sim A_2$, then

$$\mathsf{D}(\mathcal{A}_1) = \mathsf{D}(\mathcal{A}_2) = \mathsf{D}(L).$$

- $\mathsf{D}_L(x, y) = (x^{-1}L)\Delta(y^{-1}L)$.
- $\mathsf{D}(L)$ is suffix closed.
-

$$\mathsf{D}(L) = \mathsf{suff}(L) \cap \mathsf{suff}(\overline{L}).$$

- D always has a fixed-point and

$$\mathsf{D}^3(L) = \mathsf{D}^2(L).$$

Examples

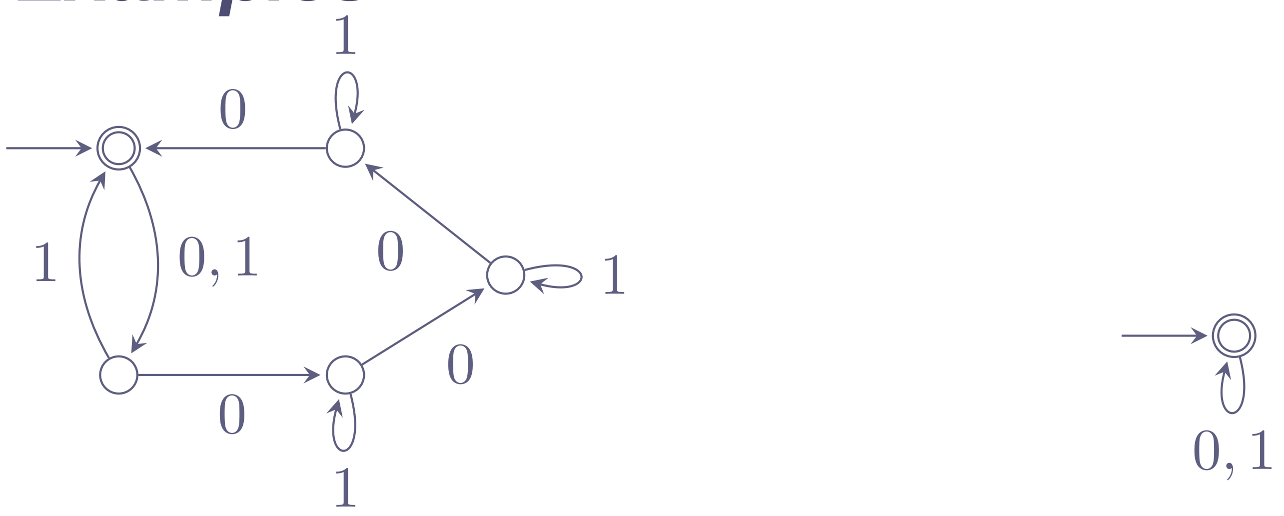

Figure 2: Automaton A_5 (left) and its distinguishability language, $\Sigma^\star$ (right).

Examples 1

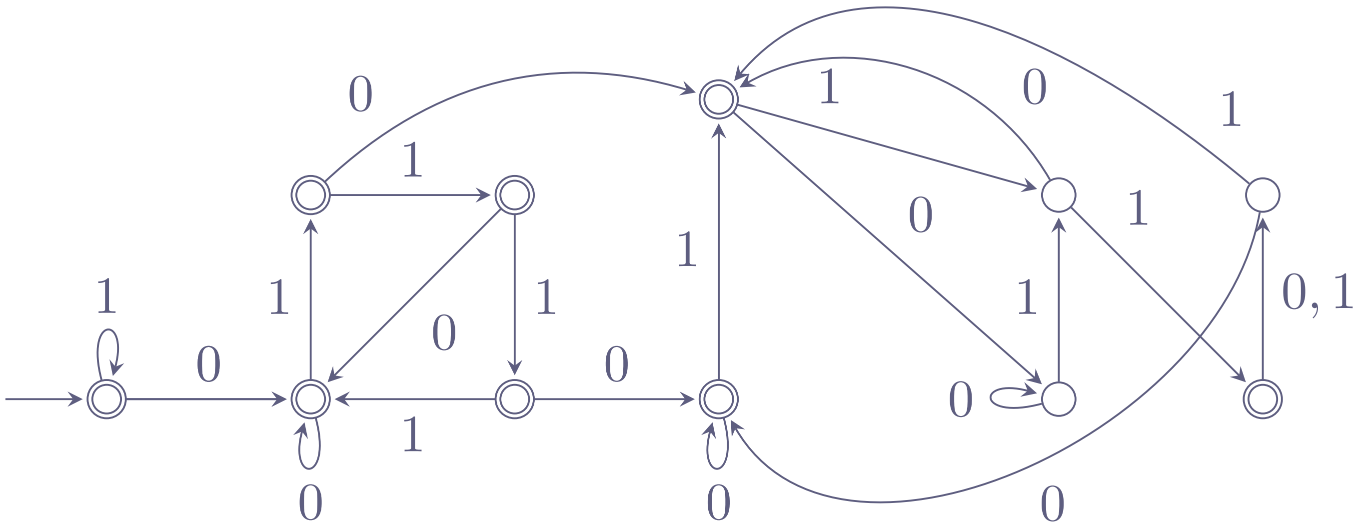

Figure 3: Example of an automaton with $\mathcal{L}(\mathcal{A}) \neq \mathrm{D}(\mathcal{L}(\mathcal{A}))$.

Examples 2

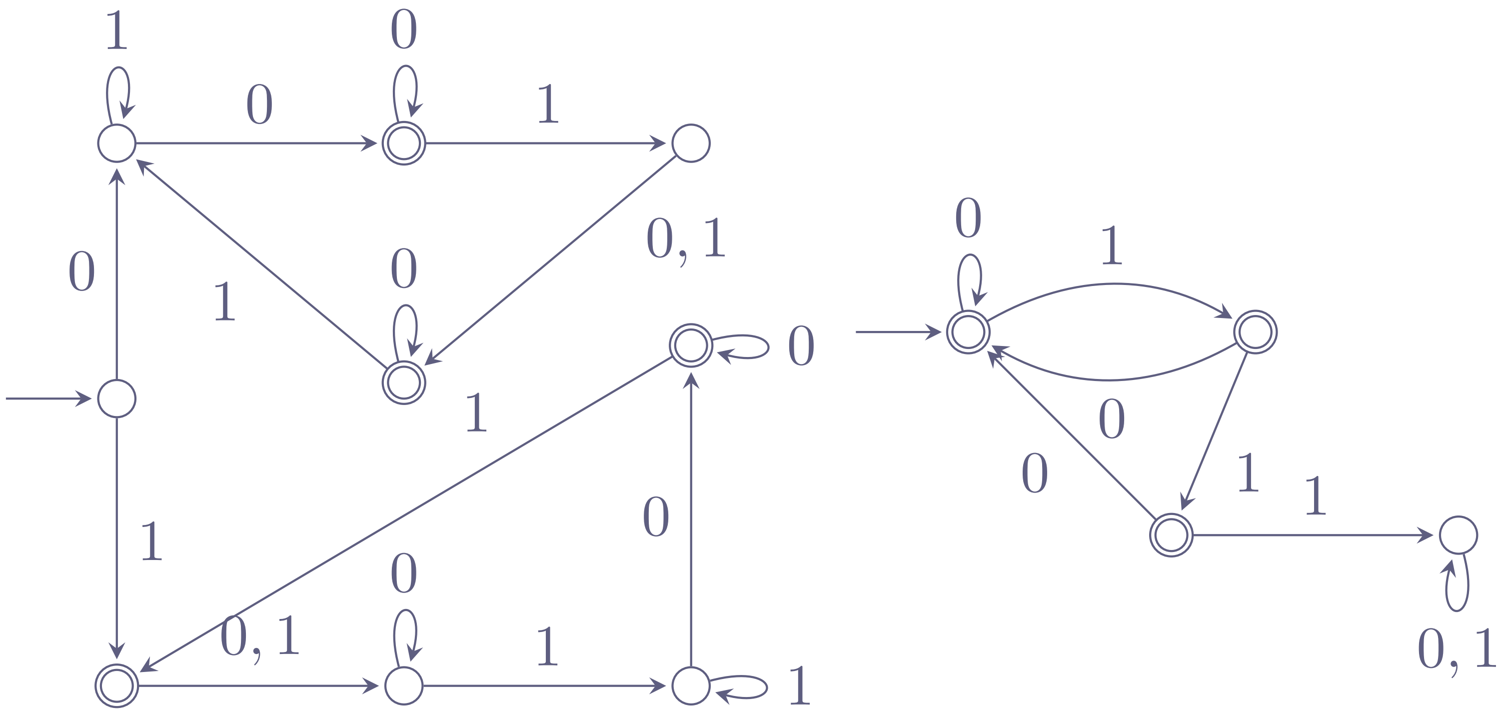

Figure 4: Example of a language L with $\mathrm{D}(L) \neq \mathrm{D}^2(L) = \mathrm{D}^n(L)$, for $n \geq 3$.

Examples 3

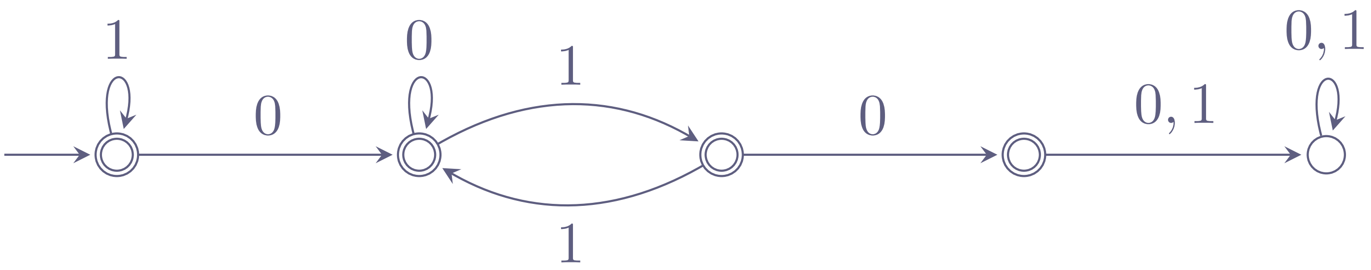

Figure 5: Example of automaton where $\mathrm{D}(\mathcal{L}(\mathcal{A}_1)) = \mathcal{L}(\mathcal{A}_1)$, i.e. distinguishability language is the same as the language of the words it can distinguish.

Results for the Combined Operations

Summary of Results

- We would like to find the upper bound for
  - $sc(\mathsf{D}(L_1 \cup L_2))$
  - $sc(\mathsf{D}(L_1) \cup \mathsf{D}(L_2))$
- We will analyze
  1. the restricted case when L_1 and L_2 have the same alphabet.
  2. the unrestricted restricted case when L_1 and L_2 have the same alphabet.
- We show the upperbound for $sc(\mathsf{D}(L_1 \odot L_2))$ in case of other Boolean operations $\odot$ for the restricted case, only.

Restricted case $sc(\mathsf{D}(L_1 \cup L_2))$

For the upper bound, if $sc(L_1) = m$ and $sc(L_2) = n$ then

- The computed theoretical upper bound (computed using the upper bound of each operation) is

$$sc(\mathsf{D}(L_1 \cup L_2)) \leq 2^{mn}$$

- The real upper bound is only

$$sc(\mathsf{D}(L_1 \cup L_2)) \leq (2^m - 1)(2^n - 1) - mn + 1$$

The Construction

$\mathcal{A}_i = \langle Q_i, \Sigma, \delta_i, 0_i, F_i \rangle$, $i = 1, 2$.
$\mathcal{A} = \langle Q, \Sigma, \delta, (0_1, 0_2), F \rangle$, $Q = Q_1 \times Q_2$,
$\delta((p, q), a) = (\delta_1(p, a), \delta_2(p, a))$, for all $a \in \Sigma^*$, and $(p, q) \in Q$,
$F = \{(p, q) \in Q \mid p \in F_1$ or $q \in F_2\}$.
For $\mathrm{D}(L)$, we have the following DFA:
$\mathcal{B} = \mathcal{B}(\mathcal{A}_1, \mathcal{A}_2) = \langle Q_{\mathcal{B}}, \Sigma, \delta_{\mathcal{B}}, Q, F_{\mathcal{B}} \rangle$, where $Q_{\mathcal{B}} \subseteq 2^Q$,
$\delta_{\mathcal{B}}(P, a) = \delta(P, a)$, and
$F_{\mathcal{B}} = \{P \in 2^{Q_1 \times Q_2} \mid$ there is $(p_1, p_2), (q_1, q_2) \in P,$ such that, $p_1 \in F_1$ or $p_2 \in F_2,$ and $(q_1 \notin F_1$ and $q_2 \notin F_2)\}$.

- We can see that the only reachable states are of the form $X_1 \times X_2$, with $X_i \subseteq Q_i$, $i = 1, 2$.
- All singleton states are equivalent to the dead state.

The Brzozowski Universal Witness

- The usual witness for many operations is given by a dialect of the Brzozowski universal witness $\mathcal{L}(D_n)$ is defined as $(\mathcal{L}(\mathcal{D}_n(a,b,c)) \mid n \geq 3)$, where $\mathcal{D}_n = \mathcal{D}_n(a,b,c) = \langle [n], \Sigma, \delta_n, 0, \{n-1\} \rangle$, $\Sigma = \{a,b,c\}$, and δ_n is defined by: δ_{na} is a circular shift $(Q+1)$, δ_{nb} is the transposition $(0,1)$, δ_{nc} is the reduction by one $((0,1) \to 0)$.

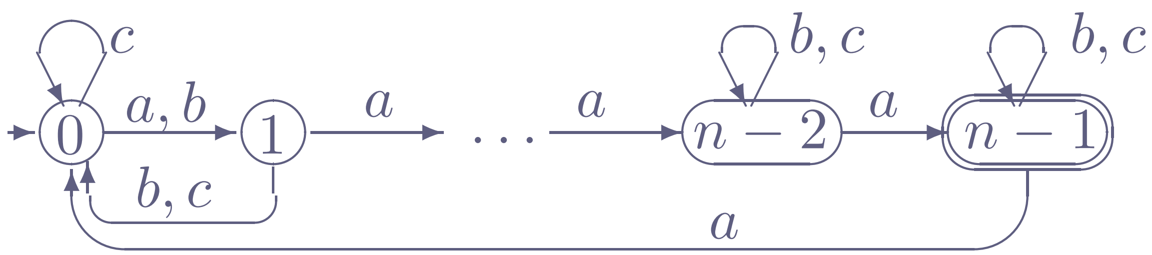

Figure 6: The universal Brzozowski witness: $\mathcal{D}_n(a,b,c)$.

The Witnesses for the Lower Bound

We give an example for $m = 3$ and $n = 4$: $\mathcal{U}_3(a, b, -, d, e)$ and $\mathcal{U}_4(d, e, -, a, b)$.

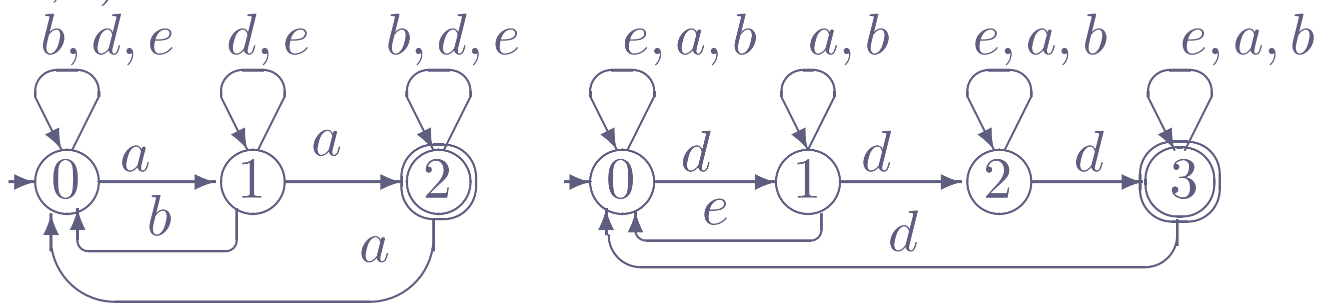

Figure 7: Example for $\mathcal{L}(\mathcal{U}_3)$ and $\mathcal{L}(\mathcal{U}_4)$, for $m = 3$ and $n = 4$, reaching the upper bound for the distinguishability of union, restricted case.

Numerical Results

		Restricted	Restricted
m	n	$\mathrm{sc}(L_m \cup L_n)$	$\mathrm{sc}(\mathsf{D}(L_m \cup L_n))$
3	3	9	41
3	4	12	94
3	5	15	203
4	4	16	210
4	5	20	446
5	5	25	937
m	n	mn	$(2^m - 1)(2^n - 1) - mn + 1$

Table 1: Numerical and theoretical results for the distinguishability of union, restricted case, confirmed by experiments.

Generalization to a Sequence of Languages

$$sc\left(\mathsf{D}\left(\bigcup_{i=1}^{r} L_i\right)\right) \leq \prod_{i=1}^{r}(2^{n_i}-1) - \prod_{i=1}^{r} n_i + 1, \qquad (1)$$

Unrestricted Case

We distinguish four cases:

- ☛ $\Sigma_1 \subset \Sigma_2$
  $\mathrm{sc}(\mathsf{D}(L_1 \cup L_2)) \leq 2^m 2^n - (2^n + m - 1)$.
- ☛ $\Sigma_2 \subset \Sigma_1$
- ☛ $\Sigma_1 \setminus \Sigma_2 \neq \emptyset$ and $\Sigma_2 \setminus \Sigma_1 \neq \emptyset$
  $\mathrm{sc}(\mathsf{D}(L_1 \cup L_2)) \leq 2^m 2^n$.
- ☛ $\Sigma_1 \cap \Sigma_2 = \emptyset$
  $\mathrm{sc}(\mathsf{D}(L_1 \cup L_2)) \leq 2^m + 2^n$.

Numerical Results

		$\Sigma_1 \subset \Sigma_2$	$\Sigma_2 \subset \Sigma_1$	$\Sigma_1 \setminus \Sigma_2 \neq \emptyset$ $\Sigma_2 \setminus \Sigma_1 \neq \emptyset$	$\Sigma_1 \cap \Sigma_2 = \emptyset$
m	n	$\mathrm{sc}(\mathsf{D}(L_m \cup L_n))$	$\mathrm{sc}(\mathsf{D}(L_m \cup L_n))$	$\mathrm{sc}(\mathsf{D}(L_m \cup L_n))$	$\mathrm{sc}(\mathsf{D}(L_m \cup L_n))$
3	3	54	54	64	16
3	4	110	117	128	24
3	5	222	244	256	40
4	4	237	237	256	32
4	5	477	492	512	48
5	5	988	988	937	64
m	n	$2^m 2^n - (2^n + m - 1)$	$2^m 2^n - (2^m + n - 1)$	$2^m 2^n$	$2^m + 2^n$

Table 2: Numerical and theoretical results for the distinguishability of union, unrestricted case, confirmed by experiments.

Distinguishability Applied to Union

Restricted Case

- $$\mathrm{sc}(\mathsf{D}(L_1) \cup \mathsf{D}(L_2)) \leq (2^m - m)(2^n - n), \tag{2}$$

- $$sc\left(\bigcup_{i=1}^{r} \mathsf{D}(L_i)\right) \leq \prod_{i=1}^{r} (2_i^n - n_i), \tag{3}$$

Unrestricted Case

- $\Sigma_1 \subset \Sigma_2$,
- $\Sigma_2 \subset \Sigma_1$,
- $\Sigma_1 \setminus \Sigma_2 \neq \emptyset$, and $\Sigma_2 \setminus \Sigma_1 \neq \emptyset$,

$$(2^m - m)(2^n - n)$$

- $\Sigma_1 \cap \Sigma_2 = \emptyset$.

$$2^m - m + 2^n - n$$

Numerical Values

		$\Sigma_2 \subset \Sigma_1$ or $\Sigma_1 \subset \Sigma_2$ or $\Sigma_1 \setminus \Sigma_2 \neq \emptyset$ $\Sigma_2 \setminus \Sigma_1 \neq \emptyset$	$\Sigma_1 \cap \Sigma_2 = \emptyset$
m	n	$\mathrm{sc}(\mathrm{D}(L_m) \cup \mathrm{D}(L_n))$	$\mathrm{sc}(\mathrm{D}(L_m) \cup \mathrm{D}(L_n))$
3	3	25	10
3	4	60	17
3	5	135	32
4	4	144	24
4	5	324	39
5	5	729	54

Table 3: State complexity of union of distinguishability, unrestricted case.

Conclusion

- The unrestricted case is only solved for the union operation.
- The complexity of Boolean operations applied to distinguishability languages is still open.
- Restrictions on the size of alphabets.
  - unary languages, i.e, included in $\{a\}^*$.
- Replace the distinguishability operation with the operation producing the set of minimal distinguishable words.
- For a language L, using Boolean operation and applying the distinguishability operation, we obtain a finite Kuratowski algebra. Find an exact bound of the state complexity for all the languages in this algebra, not just for a few of them.

Thank you!

Thank you for your attention!