Universal Disjunctive Concatenation and Star

Nelma Moreira[1] Giovanni Pighizzini[2] **Rogério Reis**[1]

[1]Centro de Matemática & Faculdade de Ciências
Universidade do Porto, Portugal

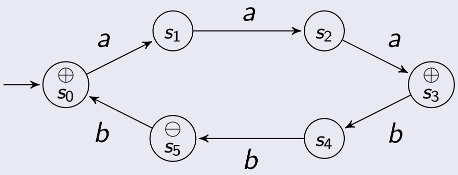
[2]Dipartimento di Informatica
Università degli Studi di Milano, Italy

DCFS 2015

Waterloo, Ontario, Canada
June 25–27, 2015

Motivation

Don't care Automata (dcDFA)



$$\mathcal{L}^{\ominus}(A),\ \mathcal{L}^{\oplus}(A)$$

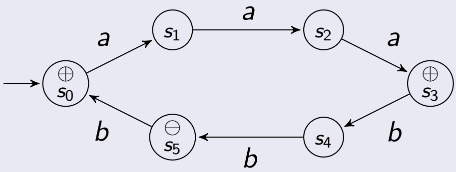
Operations over dcDFAs.

Union (disjunction):

	+	−	?
+	+	+	+
−	+	−	?
?	+	?	?

Motivation

Don't care Automata (dcDFA)



$$\mathcal{L}^{\ominus}(A),\ \mathcal{L}^{\oplus}(A)$$

Operations over dcDFAs.

For other operations, like concatenation, is much less clear, and much more interesting.

Motivation

Suppose that we have two dcDFAs A and B with completely defined behaviour, i.e.,

$$\mathcal{L}^{\oplus}(A) \cup \mathcal{L}^{\ominus}(A) = \mathcal{L}^{\oplus}(B) \cup \mathcal{L}^{\ominus}(B) = \Sigma^{\star}.$$

We want that

$$w \in \mathcal{L}^{\oplus}(AB) \Leftrightarrow \exists u, v \in \Sigma^{\star}\ w = uv \wedge u \in \mathcal{L}^{\oplus}(A) \wedge v \in \mathcal{L}^{\oplus}(B)$$

$$w \in \mathcal{L}^{\ominus}(AB) \Leftrightarrow \forall u, v \in \Sigma^{\star}\ w = uv \Rightarrow u \notin \mathcal{L}^{\oplus}(A) \vee v \notin \mathcal{L}^{\oplus}(B)$$

$$w \in \mathcal{L}^{\ominus}(AB) \Leftrightarrow \forall u, v \in \Sigma^{\star}\ w = uv \Rightarrow u \in \mathcal{L}^{\ominus}(A) \vee v \in \mathcal{L}^{\ominus}(B)$$

In general, for two dcDFAs A and B, we want to define

$$\mathcal{L}^{\oplus}(AB) = \mathcal{L}^{\oplus}(A)\mathcal{L}^{\oplus}(B) \quad \mathcal{L}^{\ominus}(AB) = \mathcal{L}^{\ominus}(A) \circledcirc \mathcal{L}^{\ominus}(B)$$

Universal Disjunctive Concatenation ($\circledcirc$)

Let L_1 and L_2 be languages over an alphabet Σ.

$$L_1 \circledcirc L_2 = \{w \in \Sigma^\star \mid \forall x_1, x_2, w = x_1 x_2 \ \Rightarrow \ x_1 \in L_1 \vee x_2 \in L_2\}$$

Example 1

$$\mathcal{L}((ab^\star a + b)^\star) \circledcirc \mathcal{L}(a(a+b)^\star) = \mathcal{L}((aa+b)^\star)$$

Example 2

$$\Sigma = \{a\} \qquad \mathcal{L}(a^\star) \circledcirc \mathcal{L}(a^\star) = \mathcal{L}(a^\star)$$

$$\Sigma = \{a, b\} \qquad \mathcal{L}(a^\star) \circledcirc \mathcal{L}(a^\star) = \mathcal{L}(a^\star + a^\star b a^\star)$$

Theorem

The class of Regular Languages is closed under the operation $\circledcirc$.

$$\overline{L \circledcirc M} = \{w \mid \exists u, v\ w = uv \wedge u \notin L \wedge v \notin M\} = \overline{L}\,\overline{M}$$

$$L \circledcirc M = \overline{\overline{L}\,\overline{M}}$$

Some properties of $\circledcirc$:

- $(L \circledcirc M) \circledcirc N = L \circledcirc (M \circledcirc N)$ (associativity)
- $\Sigma^+ \circledcirc L = L \circledcirc \Sigma^+ = L$ (Σ^+ is the unit element)
- $(L \neq \Sigma^+ \neq M) \Rightarrow L \circledcirc M \neq \Sigma^+$ (no non-trivial inverses)
- $(\varepsilon \notin L) \Rightarrow (L \circledcirc M \subseteq M) \wedge (M \circledcirc L \subseteq M)$
- $(\emptyset \circledcirc L = \mathrm{s}(L)) \wedge (L \circledcirc \emptyset = \mathrm{p}(L))$

$\mathrm{s}(L)$ interior suffix-closed of L
$\mathrm{p}(L)$ interior prefix-closed of L

- $\mathrm{pref}(L) \subseteq L \Rightarrow L \subseteq L \circledcirc M$
- $\mathrm{pref}(L) \subseteq \mathrm{pref}(L) \circledcirc L$
- $\varepsilon \notin L \Rightarrow \mathrm{pref}(L) \circledcirc L = \mathrm{pref}(L)$
- $\mathrm{pref}_+(L) \circledcirc L = L$

- $\mathrm{suff}(L) \subseteq L \Rightarrow L \subseteq M \circledcirc L$
- $\mathrm{suff}(L) \subseteq L \circledcirc \mathrm{suff}(L)$
- $\varepsilon \notin L \Rightarrow L \circledcirc \mathrm{suff}(L) = \mathrm{suff}(L)$
- $L \circledcirc \mathrm{suff}_+(L) = L$

- $(\mathrm{pref}(L) \subseteq L) \wedge (\mathrm{suff}(M) \subseteq M) \Rightarrow LM \subseteq L \circledcirc M$

Theorem

$$(sc(M) = m \wedge sc(N) = n) \Rightarrow sc(M \circledcirc N) \leq m\, 2^{n-1}$$

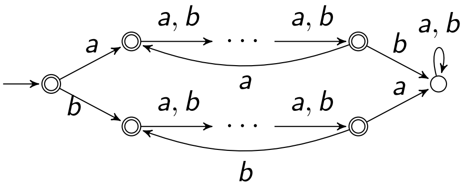
direct consequence of state complexity for concatenation [Yu, Zhuang & Salomaa, 1994]

L_n

$$(a(a+b)^{n-1})^\star(\varepsilon + a(a+b)^{<n}) + (b(a+b)^{n-1})^\star(\varepsilon + b(a+b)^{<n})$$

Theorem

- *The minimum DFA accepting L_n has $2n+2$ states*



L_n

$$(a(a+b)^{n-1})^\star(\varepsilon + a(a+b)^{<n}) + (b(a+b)^{n-1})^\star(\varepsilon + b(a+b)^{<n})$$

Theorem

- *The minimum DFA accepting L_n has $2n+2$ states*
- $\mathrm{s}(L_n) = \{x^k y \mid x \in \{a,b\}^n, k \geq 0, y \in \mathrm{pref}(x)\}$

Every non-empty word of L_n is of the form $aw_1aw_2a\cdots aw_maw'$ or $bw_1bw_2b\cdots bw_mbw'$ with $w_i \in \{a,b\}^{n-1}$ and $w' \in \{a,b\}^\star$ with $|w'| < n$.
Thus a word $w \in \mathrm{s}(L_n)$ if and only if each two symbols in w at a distance n are identical, and the proposition is trivially true.

L_n

$$(a(a+b)^{n-1})^\star(\varepsilon + a(a+b)^{<n}) + (b(a+b)^{n-1})^\star(\varepsilon + b(a+b)^{<n})$$

Theorem

- *The minimum DFA accepting L_n has $2n+2$ states*
- $\mathsf{s}(L_n) = \{x^k y \mid x \in \{a,b\}^n, k \geq 0, y \in \mathrm{pref}(x)\}$
- *Every NFA accepting $\mathsf{s}(L_n)$ has, at least, 2^n states*

Consider $S = \{x \mid x \in \{a,b\}^n\}$. We have $xx \in \mathsf{s}(L_n)$ and if $x \neq y$ then $xy \notin \mathsf{s}(L_n)$. Thus S constitutes a "fooling set" for $\mathsf{s}(L_n)$ and any NFA accepting L_n must have $|S| = 2^n$ states. □

Theorem

For every even m, there exists a minimal DFA B over a binary alphabet, with $m = |B|$, such that the minimum number of states of an NFA accepting $\emptyset \circledcirc \mathcal{L}(B)$ can be as large as $2^{(m-2)/2}$.

It is enough to consider L_n as the right operand. Given that $\emptyset \circledcirc L_n = \mathrm{s}(L_n)$ and that any NFA accepting $\mathrm{s}(L_n)$ has, at least 2^n states. Thus making $m = 2n + 2$ (the size of the smallest DFA accepting L_n) the size of the NFA accepting $\mathrm{s}(L_n)$ is, at least, $2^{(m-2)/2}$. □

Theorem

Let A' and A'' be two NFAs with m and n states $(m, n \geq 1)$, respectively. There exists an NFA A with no more than 2^{m+n} states that accepts $\mathcal{L}(A') \circledcirc \mathcal{L}(A'')$.

If $A' = \langle Q', \Sigma, \delta', i', F' \rangle$ and $A'' = \langle Q'', \Sigma, \delta'', i'', F'' \rangle$ are DFAs.

$$A = \langle Q' \times 2^{Q''}, \Sigma, \delta, i, F \rangle$$

$$i = \begin{cases} (i', \emptyset) & \text{if } i' \in F' \\ (i', \{i''\}) & \text{otherwise} \end{cases}$$

$$\delta((s, \alpha), \sigma) = \begin{cases} (\delta'(s, \sigma), \delta''(\alpha, \sigma)) & \text{if } \delta'(s, \sigma) \in F' \\ (\delta'(s, \sigma), \delta''(\alpha, \sigma) \cup \{i''\}) & \text{otherwise} \end{cases}$$

$$F = \begin{cases} F' \times 2^{F''} & \text{if } i'' \notin F'' \\ F' \times 2^{F''} \cup Q' \times 2^{F''} & \text{otherwise} \end{cases}$$

Theorem

Let A' and A'' be two NFAs with m and n states ($m, n \geq 1$), respectively. There exists an NFA A with no more than 2^{m+n} states that accepts $\mathcal{L}(A') \circledcirc \mathcal{L}(A'')$.

A is a DFA, and $\{(q, \alpha) \mid q \notin F', i'' \notin \alpha\}$ has only non-reachable states. Thus A as at most $m2^n - (m-f)2^{n-1}$ states, with $f = |F'|$.

Theorem

Let A' and A'' be two NFAs with m and n states $(m, n \geq 1)$, respectively. There exists an NFA A with no more than 2^{m+n} states that accepts $\mathcal{L}(A') \circledcirc \mathcal{L}(A'')$.

If $A' = \langle Q', \Sigma, \delta', i', F' \rangle$, a DFA, and $A'' = \langle Q'', \Sigma, \delta'', I'', F'' \rangle$, an NFA.

$$
\begin{aligned}
A &= \langle Q' \times 2^{Q''}, \Sigma, \delta, I, F \rangle \\
I &= \begin{cases} \{(i', \emptyset)\} & \text{if } i' \in F' \\ \{(i', \{i''\}) \mid i'' \in I''\} & \text{otherwise} \end{cases} \\
next(\alpha, \sigma) &= \{\gamma \in 2^{Q''} \mid \exists f : \alpha \twoheadrightarrow \gamma \text{ s.t. } (f(r) = s) \Rightarrow s \in \delta''(r, \sigma)\} \\
\delta((q, \alpha), \sigma) &= \begin{cases} \{(\delta'(q, \sigma), \beta) \mid \beta \in next(\alpha, \sigma)\} & \text{if } \delta'(q, \sigma) \in F' \\ \{(\delta(q, \sigma), \beta \cup \{i''\}) \mid i' \in I'', \beta \in next(\alpha, \sigma)\} & \text{otherwise} \end{cases} \\
F &= \begin{cases} F' \times 2^{F''} & \text{if } I'' \cap F'' = \emptyset \\ Q' \times 2^{F''} & \text{otherwise} \end{cases}
\end{aligned}
$$

Theorem

Let A' and A'' be two NFAs with m and n states $(m, n \geq 1)$, respectively. There exists an NFA A with no more than 2^{m+n} states that accepts $\mathcal{L}(A') \circledcirc \mathcal{L}(A'')$.

If both A' and A'' are NFAs, we can use the subset construction to get A' as a DFA, and proceed as in the case before. The number of states of the resulting NFA will be bounded by

$$2^{m+n}.$$

□

Universal Disjunctive Star

Let $L \subseteq \Sigma^\star$, define $L^{\circledcirc 0} = \Sigma^+$, and

$$L^{\circledcirc k} = L^{\circledcirc k-1} \circledcirc L, \quad k > 0$$

$$L^{\circledast} = \bigcap_{k \geq 0} L^{\circledcirc k}$$

$$w \in L^{\circledast} \Leftrightarrow \forall k > 0, \forall \{w_i \mid i \leq k\}, (w = w_1 w_2 \cdots w_k) \Rightarrow \exists j\, w_j \in L$$

Properties of $\circledast$:

- $\varepsilon \in L \ \Rightarrow \forall i \geq 0,\ \Sigma^{<i} \subseteq L^{\circledcirc i}$
- $w \in \Sigma^{i},\ w \in L^{\circledcirc i} \Rightarrow \forall j > i,\ w \in L^{\circledcirc j}$
- $w \in \Sigma^{i} \wedge \varepsilon \notin L, w \in L^{\circledcirc i} \Leftrightarrow \forall j > i,\ w \in L^{\circledcirc j}$
- $w \in L^{\circledast} \Leftrightarrow \forall 0 \leq i \leq |w|,\ w \in L^{\circledcirc i}$
- $L^{\circledast} = \overline{(\overline{L})^{\star}}$
- $L^{\circledast} = (L \setminus \{\varepsilon\})^{\circledast}$

Theorem

$$sc(L^{\circledast}) \leq 2^{n-1} + 2^{n-2}$$

The bound is tight.

Theorem

Given an NFA A, with $|A| = n \geq 1$, there exists an NFA A', with no more than 2^n states, such that $\mathcal{L}(A') = \mathcal{L}(A)^{\circledast}$.

First let us suppose that $A = \langle Q, \Sigma, \delta, i, F \rangle$ is a DFA and that $i \notin F$.

$$
\begin{aligned}
A' &= \langle 2^Q, \Sigma, \delta', \{i\}, 2^F \rangle \\
\delta'(\alpha, \sigma) &= \begin{cases} \delta(\alpha, \sigma) & \text{if } \delta(\alpha, \sigma) \subseteq F, \\ \delta(\alpha, \sigma) \cup \{i\} & \text{otherwise} \end{cases}
\end{aligned}
$$

To test if $w \in L^{\circledast}$, automata A needs to check that for every non null factorisation $\{w_i \mid 0 \leq i \leq h\}$, $w = w_0 w_1 \cdots w_h$ there exists a $w_i \in L$. A has already read $u = u_0 u_1 \cdots u_l$ and next symbol is σ

i) $u_0, u_1, \ldots, u_l\sigma$

ii) $u_0, u_1, \ldots, u_l, \sigma$

$\delta(i, u_l) \in F$ then all factorisations of w, starting with $u_0, u_1, \ldots, u_l$ will have a factor successfully in L. □

Theorem

Given an NFA A, with $|A| = n \geq 1$, there exists an NFA A', with no more than 2^n states, such that $\mathcal{L}(A') = \mathcal{L}(A)^{\circledast}$.

Suppose $A = \langle Q, \Sigma, \delta, \{i\}, F \rangle$, an NFA, and $i \notin F$

$$\begin{aligned}
A' &= \langle 2^Q, \Sigma, \delta', \{i\}, 2^F \rangle \\
next(\alpha, \sigma) &= \{\gamma \mid \exists f : \alpha \to 2^{\gamma}, \gamma = \bigcup_{r \in \alpha} f(r) \wedge \\
&\qquad \wedge \forall r \in \alpha (\emptyset \neq f(r) \subseteq \delta(r, \sigma))\} \\
\delta'(\alpha, \sigma) &= \{\gamma \mid \gamma \in next(\alpha, \sigma) \wedge \gamma \subseteq F\} \cup \\
&\qquad \cup \{\gamma \cup \{i\} \mid \gamma \in next(\alpha, \sigma) \wedge \gamma \not\subseteq F\}
\end{aligned}$$

□

Theorem

Given an NFA A, with $|A| = n \geq 1$, there exists an NFA A', with no more than 2^n states, such that $\mathcal{L}(A') = \mathcal{L}(A)^{\circledast}$.

Given an NFA with n states, accepting the language L, by adding one single state we can transform it into this special form, with just one initial (non-final) state, accepting $L \setminus \{\varepsilon\}$. Thus $|A'| \leq 2^{n+1}$.
But, in the automaton construction, no states $\alpha \in 2^Q$ s.t. $i \in \alpha \neq \{i\}$ are reachable, and $\emptyset$ is not reachable as well. Thus

$$|A'| \leq 2^n$$

□

Conclusions (I)

- $sc(A \circledcirc B) \leq m\, 2^{n-1}$ (this bound is tight)
- $nsc(A \circledcirc B) \leq 2^{m+n}$ (there is a witness that shows that the exponential *nsc* is unavoidable)
- $sc(A^{\circledast}) \leq 2^{n-1} + 2^{n-2}$
- $nsc(A^{\circledast}) \leq 2^{n}$

Conclusions (II)

For the non-deterministic state complexity of operations over dcNFA:

- Complementation is trivial
- For disjunction $C = A \cup B$, we consider $C^{\oplus} = A^{\oplus} \cup B^{\oplus}$, and $C^{\ominus} = A^{\ominus} \cap B^{\ominus}$. Thus C is polynomial w.r.t. $|A|$ and $|B|$.
- For concatenation $C = AB$, $C^{\oplus}$ can be build with a size $|A| + |B|$, but because $C^{\ominus} = A^{\ominus} \circledcirc B^{\ominus}$, the exponential blow is unavoidable.
- The same happens for the Kleene star $C = A^{\star}$, where $C^{\oplus} = (A^{\oplus})^{\star}$ and $C^{\ominus} = (A^{\ominus})^{\circledast}$.

Thank You!