Partial Derivative Automaton for Regular Expressions with Shuffle

Sabine Broda António Machiavelo **Nelma Moreira**
Rogério Reis

CMUP & DM-DCC,
Faculdade de Ciências da Universidade do Porto, Portugal

DCFS 2015

Waterloo, Ontario, Canada
June 25–27, 2015

Shuffle Operation

$x, y \in \Sigma^\star$ and $a, b \in \Sigma$

$$\begin{aligned} x \sqcup\!\sqcup\, \varepsilon = \varepsilon \sqcup\!\sqcup\, x &= \{x\} \\ ax \sqcup\!\sqcup\, by &= \{\, az \mid z \in x \sqcup\!\sqcup\, by \,\} \cup \{\, bz \mid z \in ax \sqcup\!\sqcup\, y \,\}. \end{aligned}$$

$L \subseteq \Sigma^\star$

$$L_1 \sqcup\!\sqcup\, L_2 = \bigcup_{x \in L_1, y \in L_2} x \sqcup\!\sqcup\, y$$

If two languages $L_1, L_2 \subseteq \Sigma^\star$ are regular then $L_1 \sqcup\!\sqcup\, L_2$ is regular.

Regular Expressions with Shuffle (RE(⧢))

$$\begin{array}{lcl} \tau & \rightarrow & \emptyset \mid \alpha \\ \alpha & \rightarrow & \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star \quad (a \in \Sigma). \end{array}$$

$$\begin{array}{lcl} \mathcal{L}(\emptyset) & = & \emptyset \\ \mathcal{L}(\varepsilon) & = & \{\varepsilon\} \\ \mathcal{L}(a) & = & \{a\} \\ \mathcal{L}(\alpha^\star) & = & \mathcal{L}(\alpha)^\star \end{array}$$

$$\begin{array}{lcl} \mathcal{L}(\alpha + \beta) & = & \mathcal{L}(\alpha) \cup \mathcal{L}(\beta) \\ \mathcal{L}(\alpha\beta) & = & \mathcal{L}(\alpha)\mathcal{L}(\beta) \\ \mathcal{L}(\alpha \sqcup\!\sqcup \beta) & = & \mathcal{L}(\alpha) \sqcup\!\sqcup \mathcal{L}(\beta). \end{array}$$

$$\varepsilon(L) = \begin{cases} \varepsilon & \text{if } \varepsilon \in L \\ \emptyset & \text{otherwise} \end{cases}$$

$$\varepsilon(\tau) = \varepsilon(\mathcal{L}(\tau))$$

Example I

$$P_3 = a_1 \sqcup\!\sqcup a_2 \sqcup\!\sqcup a_3$$

Example I

$$P_3 = a_1 \sqcup\!\sqcup\, a_2 \sqcup\!\sqcup\, a_3$$

Equivalent to

$$a_1 a_2 a_3 + a_1 a_3 a_2 + a_2 a_1 a_3 + a_2 a_3 a_1 + a_3 a_1 a_2 + a_3 a_2 a_1$$

Example I

$$P_n = a_1 \sqcup\!\sqcup \cdots \sqcup\!\sqcup a_n$$

where $n \geq 1$, $a_i \neq a_j$ for $1 \leq i \neq j \leq n$

$$\mathcal{L}(P_n) = \{\ a_{i_1} \cdots a_{i_n} \mid i_1, \ldots, i_n \text{ is a permutation of } 1, \ldots, n\}.$$

Complexity of RE($\sqcup\!\sqcup$)

Complexity of RE($\sqcup$)

- Membership for RE($\sqcup$) is NP-complete [MS94]
  (NL-complete for RE)

Complexity of RE($\sqcup\!\sqcup$)

- Membership for RE($\sqcup\!\sqcup$) is NP-complete [MS94]
  (NL-complete for RE)
- Inequivalence for RE($\sqcup\!\sqcup$) is EXP-complete [MS94]
  (PSPACE-complete for RE)

Complexity of RE($\sqcup\!\sqcup$)

- Membership for RE($\sqcup\!\sqcup$) is NP-complete [MS94]
  (NL-complete for RE)
- Inequivalence for RE($\sqcup\!\sqcup$) is EXP-complete [MS94]
  (PSPACE-complete for RE)
- RE($\sqcup\!\sqcup$) $\Longrightarrow$ NFA exponential trade-off [MS94]
  (atmost quadratic for RE)

Complexity of RE($\sqcup\!\sqcup$)

- Membership for RE($\sqcup\!\sqcup$) is NP-complete [MS94]
  (NL-complete for RE)
- Inequivalence for RE($\sqcup\!\sqcup$) is EXP-complete [MS94]
  (PSPACE-complete for RE)
- RE($\sqcup\!\sqcup$) $\Longrightarrow$ NFA exponential trade-off [MS94]
  (atmost quadratic for RE)
- RE($\sqcup\!\sqcup$) $\Longrightarrow$ RE double exponential trade-off [Gel10, GH09]

Complexity of RE($\sqcup$)

- Membership for RE($\sqcup$) is NP-complete [MS94]
  (NL-complete for RE)
- Inequivalence for RE($\sqcup$) is EXP-complete [MS94]
  (PSPACE-complete for RE)
- RE($\sqcup$) $\Longrightarrow$ NFA exponential trade-off [MS94]
  (atmost quadratic for RE)
- RE($\sqcup$) $\Longrightarrow$ RE double exponential trade-off [Gel10, GH09]
- RE($\sqcup$) $\Longrightarrow$ DFA double exponential trade-off [Gel10]
  (exponential for RE)

Automata and System of Equations

n-state NFA

$$\mathcal{A} = \langle [1,n], \{a_1, \ldots, a_k\}, S_0, \delta, F \rangle$$

right language of state i

$$\mathcal{L}_i = \{ w \mid \delta(i,w) \cap F \neq \emptyset \}$$

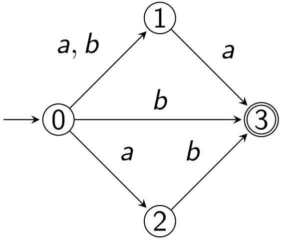
$\mathcal{L}_1, \ldots, \mathcal{L}_n$ satisfy the system of equations

$$\mathcal{L}_i \quad = \quad a_1 \mathcal{L}_{1i} \cup \cdots \cup a_k \mathcal{L}_{ki} \cup \varepsilon(\mathcal{L}_i), \quad i \in [1,n]$$

where each $\mathcal{L}_{ij}$ is a (possibly empty) union of $\mathcal{L}_m$, $m \in [1,n]$

$$\mathcal{L}(\mathcal{A}) = \bigcup_{i \in S_0} \mathcal{L}_i$$

NFA and System of Equations



$$
\begin{aligned}
\mathcal{L}_0 &= a(\mathcal{L}_1 \cup \mathcal{L}_2) \cup b(\mathcal{L}_1 \cup \mathcal{L}_3) \\
\mathcal{L}_1 &= a\ \mathcal{L}_3 \\
\mathcal{L}_2 &= b\ \mathcal{L}_3 \\
\mathcal{L}_3 &= \{\varepsilon\}
\end{aligned}
$$

$$\mathcal{L}_0 = \{b, aa, ab, ba\}$$

Regular Expressions RE($\sqcup\!\sqcup$) and System of Equations

Given $\alpha_0 \in$ RE($\sqcup\!\sqcup$), a support is set $\{\alpha_1, \ldots, \alpha_n\}$ that satisfies a system of equations

$$\alpha_i \quad = \quad a_1\alpha_{1i} + \cdots + a_k\alpha_{ki} + \varepsilon(\alpha_i), \quad i \in [0, n]$$

where α_{li}, $l \in [1, k]$, is a (possibly empty) sum of elements in $\{\alpha_1, \ldots, \alpha_n\}$.

Regular Expressions RE(⧢) and System of Equations

Given $\alpha_0 \in \mathrm{RE}(\sqcup\!\sqcup)$, a support is set $\{\alpha_1, \ldots, \alpha_n\}$ that satisfies a system of equations

$$\alpha_i \quad = \quad a_1\alpha_{1i} + \cdots + a_k\alpha_{ki} + \varepsilon(\alpha_i), \quad i \in [0, n]$$

where α_{li}, $l \in [1, k]$, is a (possibly empty) sum of elements in $\{\alpha_1, \ldots, \alpha_n\}$.

The existence of a support of α implies the existence of an NFA that accepts $\mathcal{L}(\alpha)$.

Support π

Proposition 1

Given $\tau \in \mathsf{RE}(\sqcup\!\sqcup)$, the set $\pi(\tau)$ is a support:

$$
\begin{array}{rclrcl}
\pi(\emptyset) & = & \pi(\varepsilon) = \emptyset & \pi(\alpha + \beta) & = & \pi(\alpha) \cup \pi(\beta) \\
\pi(a) & = & \{\varepsilon\} & \pi(\alpha\beta) & = & \pi(\alpha)\beta \cup \pi(\beta) \\
\pi(\alpha^\star) & = & \pi(\alpha)\alpha^\star & \pi(\alpha \sqcup\!\sqcup \beta) & = & \pi(\alpha) \sqcup\!\sqcup \pi(\beta) \\
 & & & & \cup & \pi(\alpha) \sqcup\!\sqcup \{\beta\} \cup \{\alpha\} \sqcup\!\sqcup \pi(\beta)
\end{array}
$$

Support π

Proposition 1

Given $\tau \in \mathrm{RE}(\sqcup\!\sqcup)$, the set $\pi(\tau)$ is a support:

$$
\begin{array}{rclrcl}
\pi(\emptyset) & = & \pi(\varepsilon) = \emptyset & \pi(\alpha+\beta) & = & \pi(\alpha) \cup \pi(\beta) \\
\pi(a) & = & \{\varepsilon\} & \pi(\alpha\beta) & = & \pi(\alpha)\beta \cup \pi(\beta) \\
\pi(\alpha^\star) & = & \pi(\alpha)\alpha^\star & \pi(\alpha \sqcup\!\sqcup \beta) & = & \pi(\alpha) \sqcup\!\sqcup \pi(\beta) \\
 & & & & \cup & \pi(\alpha) \sqcup\!\sqcup \{\beta\} \cup \{\alpha\} \sqcup\!\sqcup \pi(\beta)
\end{array}
$$

where, for $S, T \subseteq \mathrm{RE}(\sqcup\!\sqcup)$ and $\beta \in \mathrm{RE}(\sqcup\!\sqcup) \setminus \{\emptyset, \varepsilon\}$,

$S\beta = \{ \alpha\beta \mid \alpha \in S \}$
$S \sqcup\!\sqcup T = \{ \alpha \sqcup\!\sqcup \beta \mid \alpha \in S, \beta \in T \}$
$S\varepsilon = \{\varepsilon\} \sqcup\!\sqcup S = S \sqcup\!\sqcup \{\varepsilon\} = S$
$S\emptyset = \emptyset S = \emptyset$.

Support π

Proposition 1

Given $\tau \in \mathsf{RE}(\sqcup\!\sqcup)$, *the set* $\pi(\tau)$ *is a support:*

$$\begin{array}{rcl} \pi(\emptyset) & = & \pi(\varepsilon) = \emptyset \\ \pi(a) & = & \{\varepsilon\} \\ \pi(\alpha^\star) & = & \pi(\alpha)\alpha^\star \end{array} \qquad \begin{array}{rcl} \pi(\alpha+\beta) & = & \pi(\alpha)\cup\pi(\beta) \\ \pi(\alpha\beta) & = & \pi(\alpha)\beta\cup\pi(\beta) \\ \pi(\alpha \sqcup\!\sqcup \beta) & = & \pi(\alpha) \sqcup\!\sqcup \pi(\beta) \\ & \cup & \pi(\alpha) \sqcup\!\sqcup \{\beta\} \cup \{\alpha\} \sqcup\!\sqcup \pi(\beta) \end{array}$$

Proposition 2

$|\pi(\tau)| \leq 2^{|\tau|_\Sigma} - 1$, *where* $|\tau|_\Sigma$ *denotes the number of alphabet symbols in* τ *(= alphabetic length).*

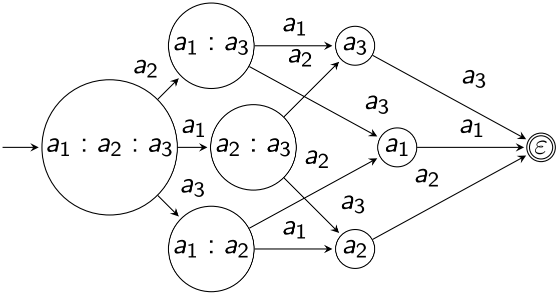
Example I

$P_3 = a_1 \sqcup\!\sqcup a_2 \sqcup\!\sqcup a_3$

Example I

$P_3 = a_1 \sqcup\!\sqcup\, a_2 \sqcup\!\sqcup\, a_3$

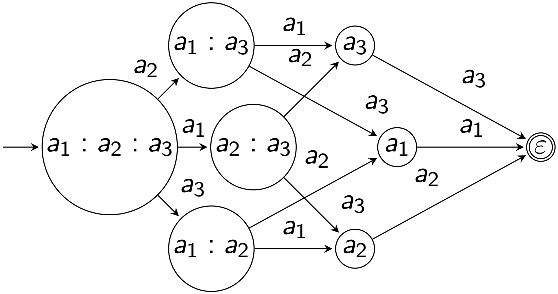
$$\pi(P_3) = \{\varepsilon, a_1, a_2, a_3, a_1 \sqcup\!\sqcup\, a_2, a_1 \sqcup\!\sqcup\, a_3, a_2 \sqcup\!\sqcup\, a_3\}$$



Example I

$P_3 = a_1 \sqcup\!\sqcup\, a_2 \sqcup\!\sqcup\, a_3$

$$\pi(P_3) = \{\varepsilon, a_1, a_2, a_3, a_1 \sqcup\!\sqcup\, a_2, a_1 \sqcup\!\sqcup\, a_3, a_2 \sqcup\!\sqcup\, a_3\}$$



Proposition 3

$|\pi(P_n)| = 2^n - 1$ *and* $|P_n|_\Sigma = n$

Partial Derivatives

$\tau \in \mathsf{RE}(\sqcup\!\sqcup)$, $a \in \Sigma$

$$\begin{array}{rcl}
\partial_a(\emptyset) & = & \partial_a(\varepsilon) = \emptyset \\
\partial_a(b) & = & \begin{cases} \{\varepsilon\} \text{ if } b = a \\ \emptyset \text{ otherwise} \end{cases} \\
\partial_a(\alpha^\star) & = & \partial_a(\alpha)\alpha^\star \\
\partial_a(\alpha + \beta) & = & \partial_a(\alpha) \cup \partial_a(\beta) \\
\partial_a(\alpha\beta) & = & \partial_a(\alpha)\beta \ \cup \ \varepsilon(\alpha)\partial_a(\beta) \\
\partial_a(\alpha \sqcup\!\sqcup \beta) & = & \partial_a(\alpha) \sqcup\!\sqcup \{\beta\} \cup \{\alpha\} \sqcup\!\sqcup \partial_a(\beta)
\end{array}$$

Partial Derivatives

$\tau \in \mathrm{RE}(\sqcup\!\sqcup),\ a \in \Sigma$

$$
\begin{array}{rcl}
\partial_a(\emptyset) & = & \partial_a(\varepsilon) = \emptyset \\
\partial_a(b) & = & \begin{cases} \{\varepsilon\} \text{ if } b = a \\ \emptyset \text{ otherwise} \end{cases} \\
\partial_a(\alpha^\star) & = & \partial_a(\alpha)\alpha^\star \\
\partial_a(\alpha + \beta) & = & \partial_a(\alpha) \cup \partial_a(\beta) \\
\partial_a(\alpha\beta) & = & \partial_a(\alpha)\beta \ \cup \ \varepsilon(\alpha)\partial_a(\beta) \\
\partial_a(\alpha \sqcup\!\sqcup \beta) & = & \partial_a(\alpha) \sqcup\!\sqcup \{\beta\} \cup \{\alpha\} \sqcup\!\sqcup \partial_a(\beta)
\end{array}
$$

$$\mathcal{L}(\partial_a(\tau)) = \{w \mid aw \in \mathcal{L}(\tau)\} = a^{-1}\mathcal{L}(\tau)$$

Partial Derivatives

$$\tau \in \mathsf{RE}(\sqcup\!\sqcup),\ a \in \Sigma,\ x \in \Sigma^\star,\ S \subseteq \mathsf{RE}(\sqcup\!\sqcup)$$

$$\begin{aligned}
\partial_\varepsilon(\tau) &= \{\tau\} \\
\partial_{xa}(\tau) &= \partial_a(\partial_x(\tau))
\end{aligned}$$

Partial Derivatives

$$\tau \in \mathsf{RE}(\sqcup\!\sqcup),\ a \in \Sigma,\ x \in \Sigma^\star,\ S \subseteq \mathsf{RE}(\sqcup\!\sqcup)$$

$$\begin{aligned} \partial_\varepsilon(\tau) &= \{\tau\} \\ \partial_{xa}(\tau) &= \partial_a(\partial_x(\tau)) \end{aligned}$$

$$\begin{aligned} \partial(\tau) &= \bigcup_{x \in \Sigma^\star} \partial_x(\tau) \\ \partial^+(\tau) &= \bigcup_{x \in \Sigma^+} \partial_x(\tau) \end{aligned}$$

Partial Derivatives

$$\tau \in \mathsf{RE}(\sqcup\!\sqcup),\ a \in \Sigma,\ x \in \Sigma^\star,\ S \subseteq \mathsf{RE}(\sqcup\!\sqcup)$$

$$\begin{aligned} \partial_\varepsilon(\tau) &= \{\tau\} \\ \partial_{xa}(\tau) &= \partial_a(\partial_x(\tau)) \end{aligned}$$

$$\begin{aligned} \partial(\tau) &= \bigcup_{x\in\Sigma^\star} \partial_x(\tau) \\ \partial^+(\tau) &= \bigcup_{x\in\Sigma^+} \partial_x(\tau) \end{aligned}$$

Proposition 4

$$\forall \tau \in \mathsf{RE}(\sqcup\!\sqcup),\ \partial^+(\tau) = \pi(\tau)$$

Partial Derivative Automaton

$\tau \in \mathsf{RE}(\sqcup\!\sqcup)$

$$\mathcal{A}_{pd}(\tau) = \langle \partial(\tau), \Sigma, \{\tau\}, \delta_\tau, F_\tau \rangle$$

where

$$\begin{array}{rcl} \delta_\tau(\gamma, a) & = & \partial_a(\gamma) \\ F_\tau & = & \{\ \gamma \in \partial(\tau) \mid \varepsilon(\gamma) = \varepsilon\ \} \end{array}$$

Partial Derivative Automaton

$\tau \in \mathsf{RE}(\sqcup\!\sqcup)$

$$\mathcal{A}_{pd}(\tau) = \langle \partial(\tau), \Sigma, \{\tau\}, \delta_\tau, F_\tau \rangle$$

where

$$\begin{array}{rcl} \delta_\tau(\gamma, a) & = & \partial_a(\gamma) \\ F_\tau & = & \{ \ \gamma \in \partial(\tau) \mid \varepsilon(\gamma) = \varepsilon \ \} \end{array}$$

Partial Derivative Automaton

$\tau \in \mathsf{RE}(\sqcup\!\sqcup)$

$$\mathcal{A}_{pd}(\tau) = \langle \partial(\tau), \Sigma, \{\tau\}, \delta_\tau, F_\tau \rangle$$

where

$$\begin{array}{rcl} \delta_\tau(\gamma, a) & = & \partial_a(\gamma) \\ F_\tau & = & \{ \, \gamma \in \partial(\tau) \mid \varepsilon(\gamma) = \varepsilon \, \} \end{array}$$

Proposition 5

$$\mathcal{L}(\mathcal{A}_{pd}(\tau)) = \mathcal{L}(\tau)$$

Example I

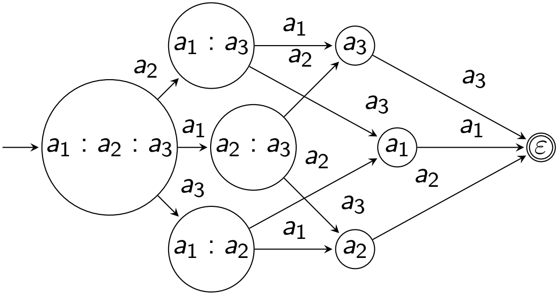
$$P_3 = a_1 \sqcup\!\sqcup\, a_2 \sqcup\!\sqcup\, a_3$$

$$\begin{array}{lcllcl}
\partial_{a_1}(P_3) & = & \{a_2 \sqcup\!\sqcup\, a_3\} & \partial_{a_1}(a_1 \sqcup\!\sqcup\, a_2) & = & \{a_2\} \\
\partial_{a_2}(P_3) & = & \{a_1 \sqcup\!\sqcup\, a_3\} & \partial_{a_2}(a_1 \sqcup\!\sqcup\, a_2) & = & \{a_1\} \\
\partial_{a_3}(P_3) & = & \{a_1 \sqcup\!\sqcup\, a_2\} & \partial_{a_2}(a_2 \sqcup\!\sqcup\, a_3) & = & \{a_3\} \\
\partial_{a_1}(a_1 \sqcup\!\sqcup\, a_3) & = & \{a_3\} & \partial_{a_3}(a_2 \sqcup\!\sqcup\, a_3) & = & \{a_3\} \\
\partial_{a_3}(a_1 \sqcup\!\sqcup\, a_3) & = & \{a_1\} & \partial_{a_i}(a_i) & = & \{\varepsilon\}
\end{array}$$

Example I

$$P_3 = a_1 \sqcup\!\sqcup\, a_2 \sqcup\!\sqcup\, a_3$$

$$\begin{array}{lcl@{\qquad}lcl}
\partial_{a_1}(P_3) & = & \{a_2 \sqcup\!\sqcup\, a_3\} & \partial_{a_1}(a_1 \sqcup\!\sqcup\, a_2) & = & \{a_2\} \\
\partial_{a_2}(P_3) & = & \{a_1 \sqcup\!\sqcup\, a_3\} & \partial_{a_2}(a_1 \sqcup\!\sqcup\, a_2) & = & \{a_1\} \\
\partial_{a_3}(P_3) & = & \{a_1 \sqcup\!\sqcup\, a_2\} & \partial_{a_2}(a_2 \sqcup\!\sqcup\, a_3) & = & \{a_3\} \\
\partial_{a_1}(a_1 \sqcup\!\sqcup\, a_3) & = & \{a_3\} & \partial_{a_3}(a_2 \sqcup\!\sqcup\, a_3) & = & \{a_3\} \\
\partial_{a_3}(a_1 \sqcup\!\sqcup\, a_3) & = & \{a_1\} & \partial_{a_i}(a_i) & = & \{\varepsilon\}
\end{array}$$



Average Case Complexity

$$\mathsf{RE}(\sqcup) \Rightarrow \mathsf{NFA}$$

Average Case Complexity

$$\tau \Rightarrow \mathcal{A}_{pd}(\tau)$$

RE $\Rightarrow$ NFA

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid \alpha + \alpha \mid \alpha \cdot \alpha \mid \alpha^\star \quad (a \in \Sigma)$$

$$|\alpha| = n \quad |\alpha|_\Sigma = m$$

RE $\Rightarrow$ NFA

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid \alpha + \alpha \mid \alpha \cdot \alpha \mid \alpha^{\star} \quad (a \in \Sigma)$$

$$|\alpha| = n \quad |\alpha|_{\Sigma} = m$$

- $\mathcal{A}_{pos}(\alpha)$ - Position Automaton [Glu61]

RE $\Rightarrow$ NFA

$$\alpha \quad \to \quad \varepsilon \mid a \mid \alpha + \alpha \mid \alpha \cdot \alpha \mid \alpha^\star \quad (a \in \Sigma)$$

$$|\alpha| = n \quad |\alpha|_\Sigma = m$$

- $\mathcal{A}_{pos}(\alpha)$ - Position Automaton [Glu61]
- $|\mathcal{A}_{pos}(\alpha)|_Q = m + 1$ and $|\mathcal{A}_{pos}(\alpha)|_\delta = \Theta(n^2)$

RE $\Rightarrow$ NFA

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid \alpha + \alpha \mid \alpha \cdot \alpha \mid \alpha^{\star} \quad (a \in \Sigma)$$

$$|\alpha| = n \quad |\alpha|_{\Sigma} = m$$

- $\mathcal{A}_{pos}(\alpha)$ - Position Automaton [Glu61]
- $|\mathcal{A}_{pos}(\alpha)|_Q = m + 1$ and $|\mathcal{A}_{pos}(\alpha)|_{\delta} = \Theta(n^2)$
- $\mathcal{A}_{pd}(\alpha)$ - Partial Derivative Automaton [Mir66, Ant96]

RE $\Rightarrow$ NFA

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid \alpha + \alpha \mid \alpha \cdot \alpha \mid \alpha^\star \quad (a \in \Sigma)$$

$$|\alpha| = n \quad |\alpha|_\Sigma = m$$

- $\mathcal{A}_{pos}(\alpha)$ - Position Automaton [Glu61]
- $|\mathcal{A}_{pos}(\alpha)|_Q = m + 1$ and $|\mathcal{A}_{pos}(\alpha)|_\delta = \Theta(n^2)$
- $\mathcal{A}_{pd}(\alpha)$ - Partial Derivative Automaton [Mir66, Ant96]
- $\mathcal{A}_{pd}$ is a quotient of $\mathcal{A}_{pos}$ [CZ02]

RE $\Rightarrow$ NFA

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid \alpha + \alpha \mid \alpha \cdot \alpha \mid \alpha^\star \quad (a \in \Sigma)$$

$$|\alpha| = n \quad |\alpha|_\Sigma = m$$

- $\mathcal{A}_{pos}(\alpha)$ - Position Automaton [Glu61]
- $|\mathcal{A}_{pos}(\alpha)|_Q = m + 1$ and $|\mathcal{A}_{pos}(\alpha)|_\delta = \Theta(n^2)$
- $\mathcal{A}_{pd}(\alpha)$ - Partial Derivative Automaton [Mir66, Ant96]
- $\mathcal{A}_{pd}$ is a quotient of $\mathcal{A}_{pos}$ [CZ02]
- $|\mathcal{A}_{pd}(\alpha)|_Q \leq m + 1$ and $|\mathcal{A}_{pd}(\alpha)|_\delta = \Theta(n^2)$

Average Complexity RE $\Rightarrow$ NFA

For the uniform distribution of $\alpha \in$ RE and asymptotically [Nic09, BMMR11, BMMR12]:

- $$|\alpha|_\Sigma \sim \frac{|\alpha|}{2}$$
- $$|\mathcal{A}_{pos}(\alpha)|_\delta = \Theta(|\alpha|)$$
- $$|\mathcal{A}_{pd}(\alpha)|_Q \sim \frac{|\mathcal{A}_{pos}(\alpha)|_Q}{2}$$
- $$|\mathcal{A}_{pd}(\alpha)|_\delta \sim \frac{|\mathcal{A}_{pos}(\alpha)|_\delta}{2}$$

How to obtain an estimate of the average complexity

$$\tau \Rightarrow \mathcal{A}_{pd}(\tau)$$

?

The Analytic Combinatorics Way

Generating Functions

C combinatorial class

$$C(z) = \sum z^{|c|} = \sum_n c_n z^n$$

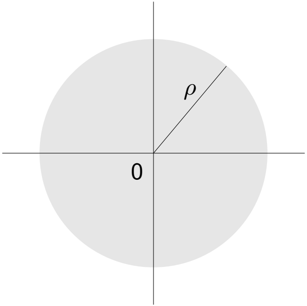
c_n: number of objects of size n

Symbolic Method

$$\begin{aligned}
\{\bullet\} &\implies U(z) = z \\
\mathsf{A} \cup \mathsf{B} &\implies A(z) + B(z) \\
\mathsf{A} \times \mathsf{B} &\implies A(z)B(z) \\
\mathsf{A}^{\star} &\implies \frac{1}{1 - A(z)}
\end{aligned}$$

Asymptotic Analysis

$$f(z) = \sum_n f_n z^n$$



$$c_n \sim \theta_n \rho^{-n}$$

where θ_n is a sub-exponential factor

Generating Function for RE(⧢) (without ∅)

$$\alpha \quad \to \quad \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star$$

$$\begin{aligned} R_k(z) &= (k+1)U(z) + \mathcal{G}(\mathcal{R}_k \times \{+\} \times \mathcal{R}_k) + \mathcal{G}(\mathcal{R}_k \times \{\cdot\} \times \mathcal{R}_k) \\ &\quad + \mathcal{G}(\mathcal{R}_k \times \{\sqcup\!\sqcup\} \times \mathcal{R}_k) + \mathcal{G}(\mathcal{R}_k \times \{^\star\}) \\ &= (k+1)U(z) + 3U(z)R_k(z)^2 + U(z)R_k(z) \\ &= (k+1)z + 3zR_k(z)^2 + zR_k(z). \end{aligned}$$

Generating Function for RE($\sqcup\!\sqcup$) (without $\emptyset$)

$$\alpha \;\to\; \varepsilon \mid a \mid (\alpha+\alpha) \mid (\alpha\cdot\alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star$$

$$\begin{aligned} R_k(z) &= (k+1)U(z) + \mathcal{G}(\mathcal{R}_k \times \{+\} \times \mathcal{R}_k) + \mathcal{G}(\mathcal{R}_k \times \{\cdot\} \times \mathcal{R}_k) \\ &\quad + \mathcal{G}(\mathcal{R}_k \times \{\sqcup\!\sqcup\} \times \mathcal{R}_k) + \mathcal{G}(\mathcal{R}_k \times \{^\star\}) \\ &= (k+1)U(z) + 3U(z)R_k(z)^2 + U(z)R_k(z) \\ &= (k+1)z + 3zR_k(z)^2 + zR_k(z). \end{aligned}$$

$$R_k(z) = \frac{(1-z)-\sqrt{\Delta_k(z)}}{6z}, \text{ where } \Delta_k(z) = 1-2z-(11+12k)z^2$$

The radius of convergence of $R_k(z)$ is $\rho_k = \frac{-1+2\sqrt{3+3k}}{11+12k}$

Generating Function for RE(⧢) (without ∅)

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star$$

$$\begin{aligned} R_k(z) &= (k+1)U(z) + \mathcal{G}(\mathcal{R}_k \times \{+\} \times \mathcal{R}_k) + \mathcal{G}(\mathcal{R}_k \times \{\cdot\} \times \mathcal{R}_k) \\ &\quad + \mathcal{G}(\mathcal{R}_k \times \{\sqcup\!\sqcup\} \times \mathcal{R}_k) + \mathcal{G}(\mathcal{R}_k \times \{^\star\}) \\ &= (k+1)U(z) + 3U(z)R_k(z)^2 + U(z)R_k(z) \\ &= (k+1)z + 3zR_k(z)^2 + zR_k(z). \end{aligned}$$

$$R_k(z) = \frac{(1-z) - \sqrt{\Delta_k(z)}}{6z}, \text{ where } \Delta_k(z) = 1 - 2z - (11+12k)z^2$$

The radius of convergence of $R_k(z)$ is $\rho_k = \frac{-1+2\sqrt{3+3k}}{11+12k}$

$$[z^n]R_k(z) \quad \sim \quad \frac{(3+3k)^{\frac{1}{4}}}{6\sqrt{\pi}} \rho_k^{-n-\frac{1}{2}} (n+1)^{-\frac{3}{2}}$$

Cumulative Generating Function for Number of Letters

$$\alpha \quad \rightarrow \quad \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star$$

$$\begin{array}{lcl@{\qquad}lcl}
l(\varepsilon) & = & 0 & l(\alpha + \beta) & = & l(\alpha) + l(\beta) \\
l(a) & = & 1 & l(\alpha \cdot \beta) & = & l(\alpha) + l(\beta) \\
l(\alpha^\star) & = & l(\alpha) & l(\alpha \sqcup\!\sqcup \beta) & = & l(\alpha) + l(\beta)
\end{array}$$

Cumulative Generating Function for Number of Letters

$$\alpha \quad \to \quad \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star$$

$$\begin{array}{lclclcl} l(\varepsilon) & = & 0 & \quad & l(\alpha+\beta) & = & l(\alpha)+l(\beta) \\ l(a) & = & 1 & & l(\alpha\cdot\beta) & = & l(\alpha)+l(\beta) \\ l(\alpha^\star) & = & l(\alpha) & & l(\alpha \sqcup\!\sqcup \beta) & = & l(\alpha)+l(\beta) \end{array}$$

$$L_k(z) \quad = \quad kz + 3zL_k(z)R_k(z) + zL_k(z)$$

$$L_k(z) = \frac{kz}{\sqrt{\Delta_k(z)}}$$

Cumulative Generating Function for Number of Letters

$$\alpha \quad \to \quad \varepsilon \mid a \mid (\alpha + \alpha) \mid (\alpha \cdot \alpha) \mid (\alpha \sqcup\!\sqcup \alpha) \mid \alpha^\star$$

$$\begin{array}{lcllcl}
l(\varepsilon) & = & 0 & l(\alpha + \beta) & = & l(\alpha) + l(\beta) \\
l(a) & = & 1 & l(\alpha \cdot \beta) & = & l(\alpha) + l(\beta) \\
l(\alpha^\star) & = & l(\alpha) & l(\alpha \sqcup\!\sqcup \beta) & = & l(\alpha) + l(\beta)
\end{array}$$

$$L_k(z) \quad = \quad kz + 3zL_k(z)R_k(z) + zL_k(z)$$

$$L_k(z) = \frac{kz}{\sqrt{\Delta_k(z)}}$$

$$[z^n]L_k(z) \quad \sim \quad \frac{k}{2\sqrt{\pi}(3+3k)^{\frac{1}{4}}} \rho_k^{-n+\frac{1}{2}} n^{-\frac{1}{2}}$$

Cumulative Generating Function for an Upper Bound of the Size of π

$$
\begin{array}{rcl rcl}
\pi(\varepsilon) & = & \emptyset & \pi(\alpha+\beta) & = & \pi(\alpha)\cup\pi(\beta) \\
\pi(a) & = & \{\varepsilon\} & \pi(\alpha\beta) & = & \pi(\alpha)\beta\cup\pi(\beta) \\
\pi(\alpha^\star) & = & \pi(\alpha)\alpha^\star & \pi(\alpha \sqcup\!\sqcup \beta) & = & \pi(\alpha)\sqcup\!\sqcup\pi(\beta) \\
 & & & & \cup & \pi(\alpha)\sqcup\!\sqcup\{\beta\}\cup\{\alpha\}\sqcup\!\sqcup\pi(\beta)
\end{array}
$$

Cumulative Generating Function for an Upper Bound of the Size of π

$$
\begin{array}{rclrcl}
\pi(\varepsilon) & = & \emptyset & \pi(\alpha+\beta) & = & \pi(\alpha)\cup\pi(\beta) \\
\pi(a) & = & \{\varepsilon\} & \pi(\alpha\beta) & = & \pi(\alpha)\beta\cup\pi(\beta) \\
\pi(\alpha^\star) & = & \pi(\alpha)\alpha^\star & \pi(\alpha \sqcup\!\sqcup \beta) & = & \pi(\alpha)\sqcup\!\sqcup\pi(\beta) \\
 & & & & \cup & \pi(\alpha)\sqcup\!\sqcup\{\beta\}\cup\{\alpha\}\sqcup\!\sqcup\pi(\beta)
\end{array}
$$

$$
\begin{array}{rclrcl}
p(\varepsilon) & = & 0 & p(\alpha+\beta) & = & p(\alpha)+p(\beta) \\
p(a) & = & 1 & p(\alpha\beta) & = & p(\alpha)+p(\beta) \\
p(\alpha^\star) & = & p(\alpha) & p(\alpha \sqcup\!\sqcup \beta) & = & p(\alpha)p(\beta)+p(\alpha)+p(\beta)
\end{array}
$$

Cumulative Generating Function for an Upper Bound of the Size of π

$$\begin{aligned} p(\varepsilon) &= 0 & p(\alpha+\beta) &= p(\alpha)+p(\beta) \\ p(a) &= 1 & p(\alpha\beta) &= p(\alpha)+p(\beta) \\ p(\alpha^\star) &= p(\alpha) & p(\alpha \sqcup\!\sqcup \beta) &= p(\alpha)p(\beta)+p(\alpha)+p(\beta) \end{aligned}$$

$$P_k(z) = kz + 6zP_k(z)R_k(z) + zP_k(z) + zP_k(z)^2$$

$$P_k(z) = \frac{\sqrt{\Delta_k(z)}}{2z} + \frac{\sqrt{\Delta'_k(z)}}{2z}$$

where $\Delta'_k(z) = 1 - 2z - (11+16k)z^2$ with zero $\rho'_k = \frac{-1+2\sqrt{3+4k}}{11+16k}$

Cumulative Generating Function for an Upper Bound of the Size of π

$$\begin{array}{rclrcl} p(\varepsilon) & = & 0 & p(\alpha+\beta) & = & p(\alpha)+p(\beta) \\ p(a) & = & 1 & p(\alpha\beta) & = & p(\alpha)+p(\beta) \\ p(\alpha^\star) & = & p(\alpha) & p(\alpha \sqcup\!\sqcup \beta) & = & p(\alpha)p(\beta)+p(\alpha)+p(\beta) \end{array}$$

$$P_k(z) \quad = \quad kz + 6zP_k(z)R_k(z) + zP_k(z) + zP_k(z)^2$$

$$P_k(z) \quad = \quad \frac{\sqrt{\Delta_k(z)}}{2z} + \frac{\sqrt{\Delta'_k(z)}}{2z}$$

where $\Delta'_k(z) = 1 - 2z - (11+16k)z^2$ with zero $\rho'_k = \frac{-1+2\sqrt{3+4k}}{11+16k}$

$$[z^n]P_k(z) \quad \sim \quad \frac{-(3+3k)^{\frac{1}{4}}\rho_k^{-n-\frac{1}{2}} + (3+4k)^{\frac{1}{4}}(\rho'_k)^{-n-\frac{1}{2}}}{2\sqrt{\pi}}(n+1)^{-\frac{3}{2}}$$

Average Size

$$avL = \frac{[z^n]L_k(z)}{[z^n]R_k(z)} = \frac{3k\rho_k}{\sqrt{3+3k}}\frac{(n+1)^{\frac{3}{2}}}{n^{\frac{1}{2}}}$$

Average Size

$$avL = \frac{[z^n]L_k(z)}{[z^n]R_k(z)} = \frac{3k\rho_k}{\sqrt{3+3k}} \frac{(n+1)^{\frac{3}{2}}}{n^{\frac{1}{2}}}$$

$$avP = \frac{[z^n]P_k(z)}{[z^n]R_k(z)}$$

Average Size

$$avL = \frac{[z^n]L_k(z)}{[z^n]R_k(z)} = \frac{3k\rho_k}{\sqrt{3+3k}} \frac{(n+1)^{\frac{3}{2}}}{n^{\frac{1}{2}}}$$

$$avP = \frac{[z^n]P_k(z)}{[z^n]R_k(z)}$$

$$\lim_{n,k\to\infty} \frac{\log_2 avP}{avL} = \log_2 \frac{4}{3} \sim 0.415$$

$$\lim_{n,k\to\infty} avP^{1/avL} = \frac{4}{3}$$

Average Size

$$avL = \frac{[z^n]L_k(z)}{[z^n]R_k(z)} = \frac{3k\rho_k}{\sqrt{3+3k}} \frac{(n+1)^{\frac{3}{2}}}{n^{\frac{1}{2}}}$$

$$avP = \frac{[z^n]P_k(z)}{[z^n]R_k(z)}$$

$$\lim_{n,k\to\infty} \frac{\log_2 avP}{avL} = \log_2 \frac{4}{3} \sim 0.415$$

$$\lim_{n,k\to\infty} avP^{1/avL} = \frac{4}{3}$$

Proposition 6

For large values of k and n an upper bound for the average number of states of $\mathcal{A}_{pd}$ is $(\frac{4}{3} + o(1))^{|\alpha|_\Sigma}$.

Final Remarks

- Experimental results suggest that the upper bound obtained may be lowered

Final Remarks

- Experimental results suggest that the upper bound obtained may be lowered
- Other shuffle operations should be analysed from an average case point of view.

Final Remarks

- Experimental results suggest that the upper bound obtained may be lowered
- Other shuffle operations should be analysed from an average case point of view.
- Similar analysis for regular expressions with intersection or complement

Thank you for your attention!

V. M. Antimirov.
Partial derivatives of regular expressions and finite automaton constructions.
Theoret. Comput. Sci., 155(2):291–319, 1996.

Sabine Broda, António Machiavelo, Nelma Moreira, and Rogério Reis.
On the average state complexity of partial derivative automata.
International Journal of Foundations of Computer Science, 22(7):1593–1606, 2011.

Sabine Broda, António Machiavelo, Nelma Moreira, and Rogério Reis.
On the average size of Glushkov and partial derivative automata.
International Journal of Foundations of Computer Science, 23(5):969–984, 2012.

J. M. Champarnaud and D. Ziadi.

Canonical derivatives, partial derivatives and finite automaton constructions.
Theoret. Comput. Sci., 289:137–163, 2002.

Wouter Gelade.
Succinctness of regular expressions with interleaving, intersection and counting.
Theor. Comput. Sci., 411(31-33):2987–2998, 2010.

Hermann Gruber and Markus Holzer.
Tight bounds on the descriptional complexity of regular expressions.
In Volker Diekert and Dirk Nowotka, editors, *13th Developments in Language Theory (DLT 2009)*, volume 5583 of *LNCS*, pages 276–287. Springer, 2009.

V. M. Glushkov.
The abstract theory of automata.
Russian Math. Surveys, 16:1–53, 1961.

B. G. Mirkin.

An algorithm for constructing a base in a language of regular expressions.
Engineering Cybernetics, 5:51—57, 1966.

Alain J. Mayer and Larry J. Stockmeyer.
Word problems-this time with interleaving.
Inf. Comput., 115(2):293–311, 1994.

Cyril Nicaud.
On the average size of Glushkov's automata.
In Adrian Horia Dediu, Armand-Mihai Ionescu, and Carlos Martín-Vide, editors, *Proc. 3rd LATA*, volume 5457 of *LNCS*, pages 626–637. Springer, 2009.