Concatenation and Kleene Star on Deterministic Finite Automata

GQ Zhang, Xiangnan Zhou, Robert Fraser, Licong Cui

Department of EECS, Case School of Engineering
Division of Medical Informatics, School of Medicine
Case Western Reserve University
Cleveland, Ohio, USA
Automata and Boolean Matrices

\[ \Delta^a = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \Delta^b = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]

\[ \Delta^\epsilon = I; \quad \Delta^{abba} = \Delta^a \Delta^b \Delta^b \Delta^b \Delta^a \]
Automata and Boolean Matrices

- Each $n$-state DFA $M = (Q, \Sigma, \delta, q_0, F)$ determines a matrix system $\{\Delta^a | a \in \Sigma\}$, where $\Delta^a$ is the $(n \times n)$ adjacency matrix of the $a$-labeled subgraph associated with the DFA.

- For a string $w = a_1 a_2 \cdots a_n$ over $\Sigma$, we write $\Delta^w$ for the matrix product $\Delta^{a_1} \Delta^{a_2} \cdots \Delta^{a_n}$. The language accepted by $M$, denoted $\mathcal{L}(M)$, is the set $\{w | I_{q_0} \Delta^w I_F^t = 1\}$.

The characteristic vector of a subset $A$ of $\{1, \cdots, n\}$ is the row vector $I^n_A \in \mathcal{B}_n$ such that the $p$-th component of $I^n_A$ is a 1 if and only if $p \in A$. The characteristic vector of a singleton set $\{p\}$ is written as $I^n_p$, or simply $I_p$. A column vector is the transpose $(\ )^t$ of a row vector.

Example: Brzozowski’s derivation

- \( u^{-1}L := \{w \mid uw \in L\} \); show \( u^{-1} \) preserves regularity
- Suppose \( L \) is accepted by an \( n \)-state DFA \( M = (Q, \Sigma, \delta, F) \), with \( \{\Delta^a \mid a \in \Sigma\} \) its matrix system.
- Then a DFA accepting \( u^{-1}L \) can be given as \( M' = (Q', \Sigma, \delta', q'_0, F') \), where
  \[
  \begin{align*}
  Q' &= \{A \mid A \in B_{n \times n}\}, \\
  q'_0 &= \Delta^u, \\
  \delta'(A, a) &= A\Delta^a, \\
  F' &= \{A \mid I_1 A I_F^t = 1\}.
  \end{align*}
  \]
- This approach can simulate inductive constructions such as Fibonacci numbers and polynomials (\( f(n) \) defined inductively in \( n \))
Concatenation

Suppose matrix systems \( \{ \Delta_1^a \mid a \in \Sigma \} \) and \( \{ \Delta_2^a \mid a \in \Sigma \} \) are associated with \( m \)- and \( n \)-state DFAs \( M_1 = (Q_1, \Sigma, \delta_1, F_1) \) and \( M_2 = (Q_2, \Sigma, \delta_2, F_2) \), respectively. The DFA \( M = (Q, \Sigma, \delta, q_0, F) \)

\[
Q = \{(A, B) \mid A \in \mathcal{B}_{m \times m}, B \in \mathcal{B}_{m \times n}\},
\]
\[
q_0 = (T^0, T),
\]
\[
\delta((A, B), a) = (A, B)\Delta^a
\]
\[
(= (A\Delta_1^a, A\Delta_1^a T + B\Delta_2^a)),
\]
\[
F = \{(A, B) \mid I_m^t B I_n^t = 1\},
\]

where \( \Delta^a = \begin{pmatrix} \Delta_1^a & \Delta_1^a T \\ 0 & \Delta_2^a \end{pmatrix} \) for \( a \in \Sigma \), \( T = I_F^t I_1^n \), and \( T^0 \) is the \((m \times m)\) identity matrix, has the property that \( \mathcal{L}(M) = \mathcal{L}(M_1) \circ \mathcal{L}(M_2) \).
Lemma for Concatenation

For the DFA $M = (Q, \Sigma, \delta, q_0, F)$

$$Q = \{(A, B) \mid A \in B_{m \times m}, B \in B_{m \times n}\},$$

$$q_0 = (T^0, T),$$

$$\delta((A, B), a) = (A, B)\Delta^a$$

$$= (A\Delta_1^a, A\Delta_1^a T + B\Delta_2^a),$$

$$F = \{(A, B) \mid I_1^m B I_{F_2}^t = 1\},$$

suppose $\delta(q_0, w) = (A, B)$ in $M$, and suppose $w = a_1 \cdots a_\ell$.

We have

$$B = \sum_{i=0}^{\ell} \Delta_1^{a_1 a_2 \cdots a_i} T \Delta_2^{a_{i+1} a_{i+2} \cdots a_\ell}.$$
Kleene Star

Suppose the matrix system \( \{ \Delta^a_i \mid a \in \Sigma \} \) is associated with an \( n \)-state DFA \( M_1 = (Q_1, \Sigma, \delta_1, F_1) \). The DFA \( M = (Q, \Sigma, \delta, q_0, F) \) with \( H = I_{F_1}^t I_1 \) and

\[
Q = \{ A \mid A \in B_{n \times n} \} \cup \{ s \},
q_0 = s,
\delta(q, a) = \begin{cases} 
\Delta^a_1 (H^0 + H^1), & \text{if } q = s, \\
A \Delta^a_1 (H^0 + H^1), & \text{if } q = A,
\end{cases}
\]

\( F = \{ A \mid I_1 A I_{F_1}^t = 1 \} \cup \{ s \}, \)

has the property that \( \mathcal{L}(M) = (\mathcal{L}(M_1))^* \). Here, \( H^1 = H \) and \( H^0 \) is the identity matrix.
Lemma for Kleene Star

For the DFA $M = (Q, \Sigma, \delta, q_0, F)$ with $H = I_{F_1}^t I_1$ and

\[
Q = \{ A \mid A \in B_{n \times n} \} \cup \{ s \},
\]
\[
q_0 = s,
\]
\[
\delta(q, a) = \begin{cases} 
\Delta_1^a(H^0 + H^1), & \text{if } q = s, \\
A\Delta_1^a(H^0 + H^1), & \text{if } q = A,
\end{cases}
\]
\[
F = \{ A \mid I_1 A I_{F_1}^t = 1 \} \cup \{ s \},
\]

we have (where $\ell = |w|$)

\[
\delta(s, w) = \sum_{w = w_1 \cdots w_k, 1 \leq k \leq \ell, w_j \neq \epsilon, 1 \leq j \leq k, i = 0, 1} \Delta_1^{w_1} H \Delta_1^{w_2} H \cdots \Delta_1^{w_k} H^i.
\]
Conclusion

- Operations on regular expressions can be directly translated to constructions on DFA.
- We obtained along the way a proof of the classical Kleene’s Theorem avoiding the use of NFA (using Arden’s Lemma in the other direction).
- Laws of Boolean matrices capture language operations inductively and algebraically.
- The “natural” constructions using matrix systems are also optimal in the usage of the number of (reachable) states, agreeing with known results in state complexity.
- See newton.case.edu/pub.html for more details