Theory of Átomata

Hellis Tamm
Institute of Cybernetics, Tallinn

Theory Seminar, April 21, 2011

Joint work with Janusz Brzozowski, accepted to DLT 2011

This work was supported by the Natural Sciences and Engineering Research Council of Canada under grant No. OGP0000871, by the Estonian Center of Excellence in Computer Science, EXCS, financed by the European Regional Development Fund, and by the Estonian Science Foundation grant 7520.

Introduction

- Nondeterministic finite automata (NFAs), introduced by Rabin and Scott in 1959, play a major role in the theory of automata and regular languages.
- For many purposes it is necessary to convert an NFA to a deterministic finite automaton (DFA).
- In particular, for every regular language there exists a unique minimal DFA.
- As well, it is possible to associate an NFA with each regular language (universal automaton, canonical residual automaton).

Our results

- We define a unique NFA — an *átomaton* — for every regular language.
- It has non-empty intersections of complemented and uncomplemented quotients — the *atoms* of the language — as its states.
- We introduce *atomic* NFAs, in which the right language of any state is a union of some atoms.
- This is a generalization of residual NFAs in which the right language of any state is a left quotient (which we prove to be a union of atoms), and includes also átomata (where the right language of any state is an atom), trim DFAs, and the trim parts of universal automata.

Main result

- We characterize the class of NFAs for which the subset construction yields a minimal DFA.
- More specifically, we show that the subset construction applied to a trim NFA produces a minimal DFA if and only if the reverse automaton of that NFA is atomic.
- This generalizes Brzozowski's method for DFA minimization by double reversal.

Automata and languages

An NFA is a quintuple $\mathcal{N} = (Q, \Sigma, \delta, I, F)$, where Q is a finite, non-empty set of *states*, Σ is a finite non-empty *alphabet*, $\delta : Q \times \Sigma \to 2^Q$ is the *transition function*, $I \subseteq Q$ is the set of *initial states*, and $F \subseteq Q$ is the set of *final states*.

The *language accepted* by an NFA $\mathcal{N}$ is
$L(\mathcal{N}) = \{w \in \Sigma^* \mid \delta(I, w) \cap F \neq \emptyset\}$.

Two NFA's are *equivalent* if they accept the same language.

The *left* and *right language* of a state q of $\mathcal{N}$ are
$L_{I,q}(\mathcal{N}) = \{w \in \Sigma^* \mid q \in \delta(I, w)\}$, and
$L_{q,F}(\mathcal{N}) = \{w \in \Sigma^* \mid \delta(q, w) \cap F \neq \emptyset\}$.

A DFA is a quintuple $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, with the transition function $\delta : Q \times \Sigma \to Q$, and the initial state q_0.

Quotients and the quotient DFA

The *left quotient* of a language L by a word w is the language $w^{-1}L = \{x \in \Sigma^* \mid wx \in L\}$.

The *quotient DFA* of a regular language L is $\mathcal{D} = (Q, \Sigma, \delta, q_0, F)$, where $Q = \{w^{-1}L \mid w \in \Sigma^*\}$, $\delta(w^{-1}L, a) = a^{-1}(w^{-1}L)$, $q_0 = \varepsilon^{-1}L = L$, and $F = \{w^{-1}L \mid \varepsilon \in w^{-1}L\}$.

Evidently, for an NFA $\mathcal{N}$, a state q of $\mathcal{N}$, and $x \in L_{I,q}(\mathcal{N})$, $L_{q,F}(\mathcal{N}) \subseteq x^{-1}(L(\mathcal{N}))$.

If $\mathcal{D}$ is a DFA and $x \in L_{q_0,q}(\mathcal{D})$, then $L_{q,F}(\mathcal{D}) = x^{-1}(L(\mathcal{D}))$.

Nondeterministic system of equations

For any language L let $L^\varepsilon = \emptyset$ if $\varepsilon \notin L$ and $L^\varepsilon = \{\varepsilon\}$ otherwise.

A *nondeterministic system of equations (NSE)* with n variables $L_1, \ldots, L_n$ is a set of language equations

$$L_i = \bigcup_{a \in \Sigma} a\Big(\bigcup_{j \in J_{i,a}} L_j\Big) \cup L_i^\varepsilon \quad i = 1, \ldots, n, \tag{1}$$

together with an *initial set of variables* $\{L_i \mid i \in I\}$, where $I, J_{i,a} \subseteq \{1, \ldots, n\}$.

The language defined by an NSE is $L = \bigcup_{i \in I} L_i$.

Each NSE defines a unique NFA $\mathcal{N}$ and *vice versa*.

States of $\mathcal{N}$ correspond to the variables L_i, there is a transition $L_i \xrightarrow{a} L_j$ in $\mathcal{N}$ if and only if $j \in J_{i,a}$, the set of initial states of $\mathcal{N}$ is $\{L_i \mid i \in I\}$, and the set of final states is $\{L_i \mid L_i^\varepsilon = \{\varepsilon\}\}$.

Deterministic system of equations

A *deterministic system of equations (DSE)* is an NSE

$$L_i = \bigcup_{a \in \Sigma} a L_{i_a} \cup L_i^{\varepsilon} \quad i = 1, \ldots, n, \tag{2}$$

where $i_a \in \{1, \ldots, n\}$, $I = \{1\}$, and the empty language $\emptyset$ is retained if it appears.

Each DSE defines a unique DFA $\mathcal{D}$ and *vice versa*.

States of $\mathcal{D}$ correspond to the variables L_i, there is a transition $L_i \xrightarrow{a} L_j$ in $\mathcal{D}$ if and only if $i_a = j$, the initial state of $\mathcal{D}$ is L_1, and the set of final states is $\{L_i \mid L_i^{\varepsilon} = \{\varepsilon\}\}$.

If $\mathcal{D}$ is minimal, its DSE constitutes its *quotient equations* where every L_i is a quotient of the initial language L_1.

Atoms

Let $L_1 = L, L_2, \ldots, L_n$ be the quotients of a regular language L.

An *atom* of L is any non-empty language of the form $A = \widetilde{L_1} \cap \widetilde{L_2} \cap \cdots \cap \widetilde{L_n}$, where $\widetilde{L_i}$ is either L_i or $\overline{L_i}$, and at least one of the L_i is not complemented ($\overline{L_1} \cap \overline{L_2} \cap \cdots \cap \overline{L_n}$ is *not* an atom).

A language has at most $2^n - 1$ atoms.

An atom is *initial* if it has L_1 (rather than $\overline{L_1}$) as a term.

An atom is *final* if and only if it contains ε.

There is exactly one final atom, the atom $\widehat{L_1} \cap \widehat{L_1} \cap \cdots \cap \widehat{L_n}$, where $\widehat{L_i} = L_i$ if $\varepsilon \in L_i$, $\widehat{L_i} = \overline{L_i}$ otherwise.

Some properties of atoms

Let $A_1, \ldots, A_m$ be the atoms of L.

The following properties hold for atoms:

- Atoms are pairwise disjoint, that is, $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, m\}$, $i \neq j$.
- The quotient $w^{-1}L$ of L by $w \in \Sigma^*$ is a (possibly empty) union of atoms.
- The quotient $w^{-1}A_i$ of A_i by $w \in \Sigma^*$ is a (possibly empty) union of atoms.

Átomaton

We use a one-to-one correspondence $A_i \leftrightarrow \mathbf{A}_i$ between atoms A_i of a language L and the states $\mathbf{A}_i$ of the NFA $\mathcal{A}$ defined below.

Let $L = L_1 \subseteq \Sigma^*$ be any regular language with the set of atoms $Q = \{A_1, \ldots, A_m\}$, initial set of atoms $I \subseteq Q$, and final atom A_m.

The *átomaton* of L is the NFA $\mathcal{A} = (\mathbf{Q}, \Sigma, \delta, \mathbf{I}, \{\mathbf{A}_m\})$, where $\mathbf{Q} = \{\mathbf{A}_i \mid A_i \in Q\}$, $\mathbf{I} = \{\mathbf{A}_i \mid A_i \in I\}$, and $\mathbf{A}_j \in \delta(\mathbf{A}_i, a)$ if and only if $aA_j \subseteq A_i$, for all $A_i, A_j \in Q$.

Example

Let L be defined by the following quotient equations:

$L_1 = aL_2 \cup bL_1, \quad L_2 = aL_3 \cup bL_1 \cup \varepsilon, \quad L_3 = aL_3 \cup bL_2.$

We find the atoms using the quotient equations:

$$
\begin{aligned}
L_1 \cap L_2 \cap L_3 \quad &= (aL_2 \cup bL_1) \cap (aL_3 \cup bL_1 \cup \varepsilon) \cap (aL_3 \cup bL_2) \\
&= (aL_2 \cap aL_3 \cap aL_3) \cup (bL_1 \cap bL_1 \cap bL_2) \\
&= a(L_2 \cap L_3) \cup b(L_1 \cap L_2) \\
&= a[(L_1 \cap L_2 \cap L_3) \cup (\overline{L_1} \cap L_2 \cap L_3)] \\
&\cup b[(L_1 \cap L_2 \cap L_3) \cup (L_1 \cap L_2 \cap \overline{L_3})], \mathit{etc.}
\end{aligned}
$$

We denote $L_i \cap L_j$ by L_{ij}, $L_i \cap \overline{L_j}$ by $L_{i\bar{j}}$, etc.

Example

Noting that $L_{1\overline{2}3}$ is empty, we have the atom equations on the right:

$$
\begin{aligned}
L_1 &= aL_2 \cup bL_1, \\
L_2 &= aL_3 \cup bL_1 \cup \varepsilon, \\
L_3 &= aL_3 \cup bL_2.
\end{aligned}
\qquad
\begin{aligned}
L_{123} &= a(L_{123} \cup L_{\overline{1}23}) \cup b(L_{123} \cup L_{12\overline{3}}), \\
L_{\overline{1}23} &= aL_{\overline{1}\,\overline{2}3}, \\
L_{12\overline{3}} &= bL_{1\overline{2}\,\overline{3}}, \\
L_{\overline{1}\,\overline{2}3} &= b(L_{\overline{1}23} \cup L_{\overline{1}2\overline{3}}), \\
L_{1\overline{2}\,\overline{3}} &= a(L_{12\overline{3}} \cup L_{\overline{1}2\overline{3}}), \\
L_{\overline{1}2\overline{3}} &= \varepsilon.
\end{aligned}
$$

Example

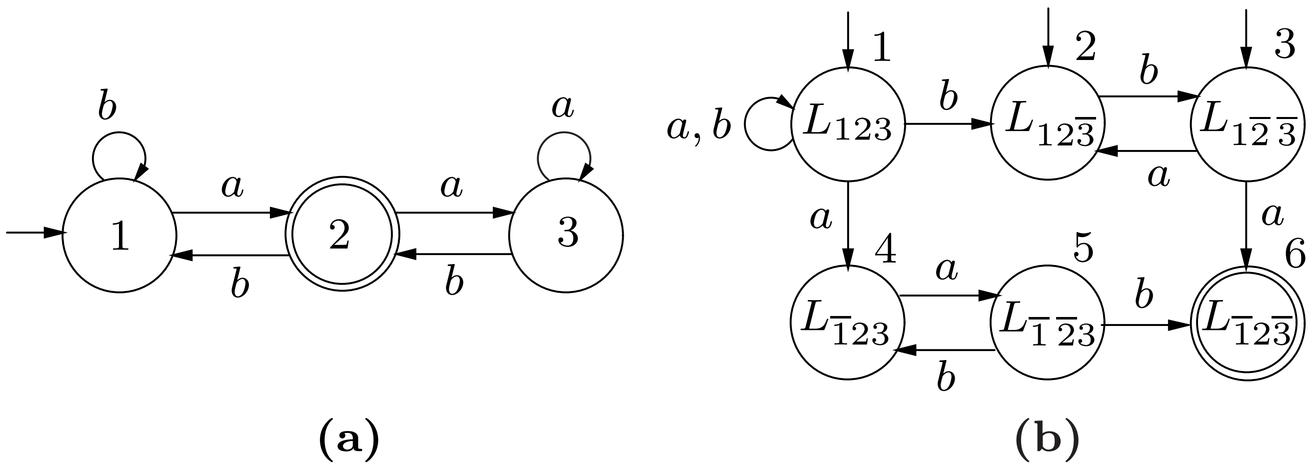

Figure 1: (a) quotient DFA; (b) átomaton

Some properties of átomaton

Let $A_1, \ldots, A_m$ be the atoms and let $\mathcal{A}$ be the átomaton of L.

- The right language of state $\mathbf{A}_i$ of $\mathcal{A}$ is the atom A_i, that is, $L_{\mathbf{A}_i,\{\mathbf{A}_m\}}(\mathcal{A}) = A_i$, for all $i \in \{1, \ldots, m\}$.
- The language accepted by $\mathcal{A}$ is L, that is, $L(\mathcal{A}) = L$.
- The reverse automaton $\mathcal{A}^{\mathbb{R}}$ of $\mathcal{A}$ is a minimal (incomplete) DFA for the reverse language of L.
- $\mathcal{A}$ is isomorphic to the minimal incomplete DFA of L if and only if L is bideterministic.

Atomic automata

We define an NFA $\mathcal{N} = (Q, \Sigma, \delta, I, F)$ to be *atomic* if for every state $q \in Q$, the right language $L_{q,F}(\mathcal{N})$ of q is a union of some atoms of $L(\mathcal{N})$.

We call an NFA $\mathcal{N}$ *residual*, if $L_{q,F}(\mathcal{N})$ is a (left) quotient of $L(\mathcal{N})$ for every $q \in Q$. Since every quotient is a union of atoms, every residual NFA is atomic.

Every trim DFA is a special case of a residual NFA; hence every trim DFA is atomic.

Naturally, the átomaton $\mathcal{A}$ is atomic since the right language of every state of $\mathcal{A}$ is an atom of L.

Also, it can be shown that the trim part of the *universal automaton* is atomic.

Extension of Brzozowski's Theorem

Theorem (Brzozowski, 1963). For a trim NFA $\mathcal{N}$, $\mathcal{N}^{\mathbb{D}}$ is minimal if $\mathcal{N}^{\mathbb{R}}$ is deterministic.

This theorem forms the basis for Brzozowski's DFA minimization algorithm: Given any DFA $\mathcal{D}$,

1) reverse it to get $\mathcal{D}^{\mathbb{R}}$,
2) determinize $\mathcal{D}^{\mathbb{R}}$ to get $\mathcal{D}^{\mathbb{R}\mathbb{D}}$,
3) reverse $\mathcal{D}^{\mathbb{R}\mathbb{D}}$ to get $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}}$,
4) determinize $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}}$ to get $\mathcal{D}^{\mathbb{R}\mathbb{D}\mathbb{R}\mathbb{D}}$.

Our generalization:

Theorem. For a trim NFA $\mathcal{N}$, $\mathcal{N}^{\mathbb{D}}$ is minimal if and only if $\mathcal{N}^{\mathbb{R}}$ is atomic.

Conclusions

- We have introduced a natural set of languages—the atoms—that are defined by every regular language.
- We defined a unique NFA for every regular language, the átomaton, and related it to other known concepts.
- We introduced atomic NFAs, and showed that some known subclasses of NFAs belong to this class.
- We characterized the class of trim NFAs for which the subset construction yields a minimal DFA.