Atoms of regular languages

Hellis Tamm
Institute of Cybernetics
Tallinn University of Technology

Brzozowski 80
Waterloo, ON, June 24, 2015

Main publications

- J. Brzozowski, H. Tamm: Theory of átomata. DLT 2011: 105-116
- J. Brzozowski, H. Tamm: Theory of átomata. Theor. Comput. Sci., 539: 13-27 (2014)
- J. Brzozowski, H. Tamm: Quotient complexities of atoms of regular languages. DLT 2012: 50-61
- J. Brzozowski, H. Tamm: Complexity of atoms of regular languages. Int. J. of Found. Comput. Sci., 24 (7): 1009-1027 (2013)
- J. Brzozowski, G. Davies: Maximally atomic languages. AFL 2014: 151-161
- J. Brzozowski, S. Davies: Quotient complexities of atoms in regular ideal languages. arXiv:1503.02208 (2015)
- S. Iván: Complexity of atoms, combinatorially. arXiv:1404.6632 (2014, 2015)

Quotients and atoms

Let L be a regular language over an alphabet Σ.

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

Quotients and atoms

Let L be a regular language over an alphabet Σ.

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

Let $K_1, \dots, K_n$ be the quotients of L.

An atom of L is any non-empty language of the form

$$A = \widetilde{K_1} \cap \widetilde{K_2} \cap \cdots \cap \widetilde{K_n},$$

where $\widetilde{K_i}$ is either K_i or $\overline{K_i}$.

Quotients and atoms

Let L be a regular language over an alphabet Σ.

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

Let $K_1, \dots, K_n$ be the quotients of L.

An atom of L is any non-empty language of the form

$$A = \widetilde{K_1} \cap \widetilde{K_2} \cap \cdots \cap \widetilde{K_n},$$

where $\widetilde{K_i}$ is either K_i or $\overline{K_i}$.

A language has at most 2^n atoms.

Atoms as congruence classes

Nerode right congruence is well known: two words x and y are equivalent if for every $v \in \Sigma^*$, $xv \in L$ if and only if $yv \in L$.

Atoms as congruence classes

Nerode right congruence is well known: two words x and y are equivalent if for every $v \in \Sigma^*$, $xv \in L$ if and only if $yv \in L$.

Myhill congruence is two-sided: the words x and y are equivalent if for all $u, v \in \Sigma^*$, $uxv \in L$ if and only if $uyv \in L$.

Atoms as congruence classes

Nerode right congruence is well known: two words x and y are equivalent if for every $v \in \Sigma^*$, $xv \in L$ if and only if $yv \in L$.

Myhill congruence is two-sided: the words x and y are equivalent if for all $u, v \in \Sigma^*$, $uxv \in L$ if and only if $uyv \in L$.

The left congruence can also be defined: two words x and y are equivalent if for every $u \in \Sigma^*$, $ux \in L$ if and only if $uy \in L$.

Equivalence classes of this relation are the atoms of L.

Atoms as congruence classes

Nerode right congruence is well known: two words x and y are equivalent if for every $v \in \Sigma^*$, $xv \in L$ if and only if $yv \in L$.

Myhill congruence is two-sided: the words x and y are equivalent if for all $u, v \in \Sigma^*$, $uxv \in L$ if and only if $uyv \in L$.

The left congruence can also be defined: two words x and y are equivalent if for every $u \in \Sigma^*$, $ux \in L$ if and only if $uy \in L$.

Equivalence classes of this relation are the atoms of L.

The view of atoms as congruence classes was first suggested by Iván (2014).

Recently, Brzozowski and Davies (2015) called this atom congruence, which is related to the Nerode and Myhill congruences in a natural way.

Properties of atoms

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

- Atoms are pairwise disjoint, that is, $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, \ldots, m\}$, $i \neq j$.
- Any quotient $w^{-1}L$ of L by $w \in \Sigma^*$ is a (possibly empty) union of atoms.
- Any quotient $w^{-1}A_i$ of an atom A_i by $w \in \Sigma^*$ is a (possibly empty) union of atoms.
- Atoms define a partition of Σ^*.
- Atoms are basic building blocks of regular languages.

Átomaton

Let $K_1 = L$ be the initial quotient of L.

- An atom is initial if it has K_1 (rather than $\overline{K_1}$) as a term.
- An atom is final if and only if it contains ε.
- There is exactly one final atom, the atom $\widehat{K_1} \cap \cdots \cap \widehat{K_n}$, where $\widehat{K_i} = K_i$ if $\varepsilon \in K_i$, $\widehat{K_i} = \overline{K_i}$ otherwise.

Átomaton

Let $K_1 = L$ be the initial quotient of L.

- An atom is initial if it has K_1 (rather than $\overline{K_1}$) as a term.
- An atom is final if and only if it contains ε.
- There is exactly one final atom, the atom $\widehat{K_1} \cap \cdots \cap \widehat{K_n}$, where $\widehat{K_i} = K_i$ if $\varepsilon \in K_i$, $\widehat{K_i} = \overline{K_i}$ otherwise.

Let $Q = \{A_1, \ldots, A_m\}$ be the set of atoms of L, with the subset $I \subseteq Q$ of initial atoms, and final atom A_m.

The átomaton of L is the NFA $\mathcal{A} = (\mathbf{Q}, \Sigma, \delta, \mathbf{I}, \{\mathbf{A}_m\})$, where

- $\mathbf{Q} = \{\mathbf{A}_i \mid A_i \in Q\}$,
- $\mathbf{I} = \{\mathbf{A}_i \mid A_i \in I\}$,
- $\mathbf{A}_j \in \delta(\mathbf{A}_i, a)$ if and only if $aA_j \subseteq A_i$, for all $A_i, A_j \in Q$.

Some properties of átomaton

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

- The language accepted by $\mathcal{A}$ is L, that is, $L(\mathcal{A}) = L$.
- The right language of state $\mathbf{A}_i$ of $\mathcal{A}$ is the atom A_i.
- The reverse automaton $\mathcal{A}^R$ of $\mathcal{A}$ is a minimal DFA for the reverse language of L.
- The determinized automaton $\mathcal{A}^D$ of $\mathcal{A}$ is a minimal DFA of L.

Atomic automata

We define an NFA $\mathcal{N}$ to be atomic if for every state q of $\mathcal{N}$, the right language of q is a union of some atoms of $L(\mathcal{N})$.

Let L be a regular language. The following are atomic:

- The átomaton of L
- The minimal DFA of L
- The universal automaton of L

Brzozowski's Theorem and DFA Minimization

Theorem (Brzozowski, 1962). For an NFA $\mathcal{N}$ without empty states, if $\mathcal{N}^R$ is deterministic, then $\mathcal{N}^D$ is minimal.

Brzozowski's Theorem and DFA Minimization

Theorem (Brzozowski, 1962). For an NFA $\mathcal{N}$ without empty states, if $\mathcal{N}^R$ is deterministic, then $\mathcal{N}^D$ is minimal.

Brzozowski's (double-reversal) DFA minimization:

Given a DFA $\mathcal{D}$ of L, the minimal DFA is obtained by $\mathcal{D}^{RDRD}$.

Works also, if $\mathcal{D}$ is replaced by an NFA.

Generalization of Brzozowski's Theorem

Theorem (Brzozowski and Tamm, 2011, 2014). For any NFA $\mathcal{N}$, $\mathcal{N}^D$ is minimal if and only if $\mathcal{N}^R$ is atomic.

Generalization of Brzozowski's Theorem

Theorem (Brzozowski and Tamm, 2011, 2014). For any NFA $\mathcal{N}$, $\mathcal{N}^D$ is minimal if and only if $\mathcal{N}^R$ is atomic.

Applications:

- A polynomial double-reversal DFA minimization algorithm (Vázquez de Parga, García, and López, 2013):

  Let $\mathcal{D}$ be a DFA with no unreachable states.
  The minimal DFA is obtained by $\mathcal{D}^{RARD}$, where A is an *atomization* algorithm (produces an atomic NFA).

Generalization of Brzozowski's Theorem

Theorem (Brzozowski and Tamm, 2011, 2014). For any NFA $\mathcal{N}$, $\mathcal{N}^D$ is minimal if and only if $\mathcal{N}^R$ is atomic.

Applications:

- A polynomial double-reversal DFA minimization algorithm (Vázquez de Parga, García, and López, 2013):

  Let $\mathcal{D}$ be a DFA with no unreachable states.
  The minimal DFA is obtained by $\mathcal{D}^{RARD}$, where A is an *atomization* algorithm (produces an atomic NFA).

- García, López, and Vázquez de Parga (2015) also showed a relationship between two main approaches for DFA minimization: partitioning of the states of a DFA, and the double-reversal method.

Quotient complexity of atoms

Quotient complexity = state complexity.

Let L have n quotients, $n \geqslant 1$.

Theorem (Brzozowski and Tamm, 2012, 2013).
For $n \geqslant 1$, the quotient complexity of the atoms with 0 or n complemented quotients is less than or equal to $2^n - 1$.
For $n \geqslant 2$ and r satisfying $1 \leqslant r \leqslant n-1$, the quotient complexity of any atom of L with r complemented quotients is less than or equal to

$$f(n,r) = 1 + \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} \binom{n}{h} \binom{h}{k}.$$

Moreover, these bounds are tight.

Quotient complexity of atoms

Quotient complexity = state complexity.

Let L have n quotients, $n \geqslant 1$.

Theorem (Brzozowski and Tamm, 2012, 2013).
For $n \geqslant 1$, the quotient complexity of the atoms with 0 or n complemented quotients is less than or equal to $2^n - 1$.
For $n \geqslant 2$ and r satisfying $1 \leqslant r \leqslant n-1$, the quotient complexity of any atom of L with r complemented quotients is less than or equal to

$$f(n,r) = 1 + \sum_{k=1}^{r} \sum_{h=k+1}^{k+n-r} \binom{n}{h} \binom{h}{k}.$$

Moreover, these bounds are tight.

Another proof for these results was suggested by Iván (2014).

Quotient complexities of atoms in language classes

- Brzozowski and Davies (2014, 2015) have studied the maximal number of atoms and the maximal quotient complexities of atoms of right, left and two-sided regular ideals.
- Brzozowski and Szykuła (2015) studied the maximal number of atoms and the maximal quotient complexities of atoms of suffix-free languages.
- Diekert and Walter (2015) studied the asymptotic behaviour of the quotient complexity of atoms.

Maximally Atomic Languages

Brzozowski and Davies (2014) defined a new class of regular languages:

A language is maximally atomic if it has the maximal number of atoms, and if every atom has the maximal complexity.

Theorem (Brzozowski and Davies, 2014). Let L be a regular language with complexity $n \geqslant 3$, and let T be the transition semigroup of the minimal DFA of L. Then L is maximally atomic if and only if the subgroup of permutations in T is set-transitive and T contains a transformation of rank $n-1$.

Maximally Atomic Languages

Brzozowski and Davies (2014) defined a new class of regular languages:

A language is maximally atomic if it has the maximal number of atoms, and if every atom has the maximal complexity.

Theorem (Brzozowski and Davies, 2014). Let L be a regular language with complexity $n \geqslant 3$, and let T be the transition semigroup of the minimal DFA of L. Then L is maximally atomic if and only if the subgroup of permutations in T is set-transitive and T contains a transformation of rank $n-1$.

Another proof for this result was presented by Iván (2014).

Conclusions

- We have introduced a natural set of languages – the atoms – that are defined by every regular language, and that are the smallest building blocks of regular languages.
- We defined a unique NFA for every regular language, the átomaton, and related it to other known concepts.
- We characterized the class of NFAs for which the subset construction yields a minimal DFA.
- We have introduced a new complexity measure for regular languages: the quotient complexity of atoms.

Sto lat, Janusz!