Theoretical Computer Science 604 (2015) 81–101

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Coalgebraic constructions of canonical nondeterministic automata

Robert S.R. Myers [a], Jiří Adámek [a], Stefan Milius [b,1], Henning Urbat [a,*]

[a] Institut für Theoretische Informatik, Technische Universität Braunschweig, Germany
[b] Lehrstuhl für Theoretische Informatik, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

ARTICLE INFO

Article history:
Received 23 September 2014
Received in revised form 9 March 2015
Accepted 21 March 2015
Available online 31 March 2015

Keywords:
Non-deterministic automata
Join-semilattices
Coalgebras
Minimization

ABSTRACT

For each regular language L we describe a family of canonical nondeterministic acceptors (nfas). Their construction follows a uniform recipe: build the minimal dfa for L in a locally finite variety $\mathcal{V}$, and apply an equivalence between the category of finite $\mathcal{V}$-algebras and a suitable category of finite structured sets and relations. By instantiating this to different varieties, we recover three well-studied canonical nfas: $\mathcal{V}$ = boolean algebras yields the átomaton of Brzozowski and Tamm, $\mathcal{V}$ = semilattices yields the jiromaton of Denis, Lemay and Terlutte, and $\mathcal{V} = \mathbb{Z}_2$-vector spaces yields the minimal xor automaton of Vuillemin and Gama. Moreover, we obtain a new canonical nfa called the distromaton by taking $\mathcal{V}$ = distributive lattices. Each of these nfas is shown to be minimal relative to a suitable measure, and we derive sufficient conditions for their state-minimality. Our approach is coalgebraic, exhibiting additional structure and universal properties of the canonical nfas.

1. Introduction

One of the core topics in classical automata theory is the construction of state-minimal acceptors for a given regular language. It is well known that the difficulty of this task depends on whether one has deterministic or nondeterministic acceptors in mind. First, every regular language L is accepted by a unique minimal *deterministic* finite automaton (dfa). Following a classical construction due to Brzozowski [12], the state set Q_L of the minimal dfa consists of all left derivatives of L; see Example 2.12. For *nondeterministic* finite automata (nfas) the situation is significantly more complex: a regular language may have many non-isomorphic state-minimal nfas, and generally there is no way to identify a "canonical" one among them. However, several authors proposed nondeterministic acceptors that *are* in some sense canonical (though not necessarily state-minimal), e.g. the *átomaton* of Brzozowski and Tamm [11], the *jiromaton*[2] of Denis, Lemay and Terlutte [13], and the *minimal xor automaton* of Vuillemin and Gama [25]. In each case, the respective nfa is formed by closing the set Q_L of left derivatives under certain algebraic operations and taking a minimal set of generators as states. Specifically:

* Corresponding author.
E-mail addresses: me@robmyers.gmail.com (R.S.R. Myers), adamek@iti.cs.tu-bs.de (J. Adámek), mail@stefan-milius.eu (S. Milius), urbat@iti.cs.tu-bs.de (H. Urbat).
1 Stefan Milius acknowledges support by the Deutsche Forschungsgemeinschaft (DFG) under project MI 717/5-1.
2 In [13] the authors called their acceptor "canonical residual finite state automaton". We propose the shorter "jiromaton" because this is analogous to the átomaton terminology.

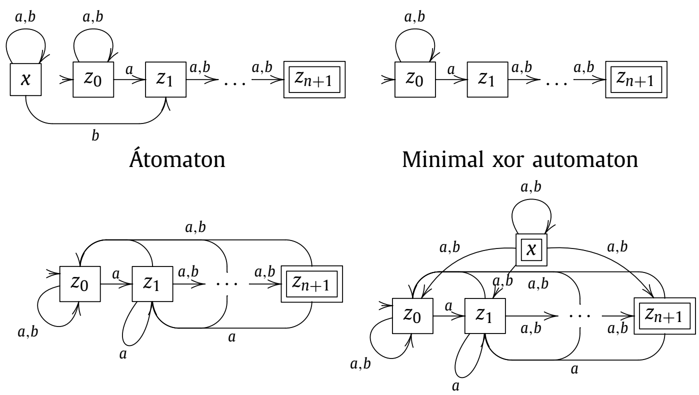
http://dx.doi.org/10.1016/j.tcs.2015.03.035
0304-3975/

1. The states of the átomaton are the atoms of the boolean algebra generated by Q_L, obtained by closing Q_L under finite union, finite intersection and complement.
2. The states of the jiromaton are the join-irreducibles of the join-semilattice generated by Q_L, obtained by closing Q_L under finite union.
3. The states of the minimal xor automaton form a basis for the $\mathbb{Z}_2$-vector space generated by Q_L. Recall that the $\mathbb{Z}_2$-vector space with basis B is the set of all finite subsets of B with $\emptyset$ as the zero vector and addition given by symmetric difference $M \oplus N = (M \setminus N) \cup (N \setminus M)$. Thus the states of the minimal xor automaton are obtained by closing Q_L under symmetric difference and choosing a basis of the resulting $\mathbb{Z}_2$-vector space.

Note that the minimal xor automaton differs substantially from the other examples treated in our paper w.r.t. the manner of language acceptance: here we consider acceptance of $\mathbb{Z}_2$-weighted languages. That is, a state accepts a word iff the number of accepting paths is odd. In the present paper we demonstrate that all these canonical nfas arise from a coalgebraic construction. For this purpose we first consider *deterministic* automata interpreted in a locally finite variety $\mathcal{V}$, where *locally finite* means that finitely generated algebras are finite. The three examples above correspond to the variety $\mathcal{V}$ of boolean algebras, join-semilattices and $\mathbb{Z}_2$-vector spaces, respectively. A *deterministic $\mathcal{V}$-automaton* is a coalgebra for the endofunctor $T_\Sigma = \mathbb{2} \times \mathsf{Id}^\Sigma$ on $\mathcal{V}$, for a fixed two-element algebra $\mathbb{2}$. In Section 2 we describe a Brzozowski-like construction that yields, for every regular language, the minimal deterministic finite $\mathcal{V}$-automaton accepting it. Next, for certain varieties $\mathcal{V}$ of interest, we derive an equivalence between the full subcategory $\mathcal{V}_f$ of finite algebras and a suitable category $\overline{\mathcal{V}}$ of finite structured sets, whose morphisms are relations preserving the structure. In each case, the objects of $\overline{\mathcal{V}}$ are "small" representations of their counterparts in $\mathcal{V}_f$, based on specific generators of algebras in $\mathcal{V}_f$. The equivalence $\mathcal{V}_f \cong \overline{\mathcal{V}}$ then induces an equivalence between deterministic finite $\mathcal{V}$-automata and coalgebras in $\overline{\mathcal{V}}$ which are *nondeterministic* automata.

Hence we have the following two-step procedure for constructing a canonical nfa for a given regular language L: (i) form the minimal deterministic $\mathcal{V}$-automaton accepting L, and (ii) use the equivalence of $\mathcal{V}_f$ and $\overline{\mathcal{V}}$ to obtain an equivalent nfa. We explain this in Section 3 and show that applying this to different varieties $\mathcal{V}$ yields the three canonical nfas mentioned above. For the átomaton one takes $\mathcal{V} = \mathsf{BA}$ (boolean algebras). Then the minimal deterministic BA-automaton for L arises from the minimal dfa by closing its states Q_L under boolean operations. The category $\overline{\mathcal{V}} = \overline{\mathsf{BA}}$ is based on Stone duality: $\overline{\mathsf{BA}}$ is the dual of the category of finite sets, so it has as objects all finite sets and as morphisms all converse-functional relations. The equivalence functor $\mathsf{BA}_f \xrightarrow{\cong} \overline{\mathsf{BA}}$ maps each finite boolean algebra to the set of its atoms. This equivalence applied to the minimal deterministic BA-automaton for L gives precisely the átomaton. Similarly, by taking $\mathcal{V}$ = join-semilattices and $\mathcal{V}$ = vector spaces over $\mathbb{Z}_2$ and describing a suitable equivalence $\mathcal{V}_f \cong \overline{\mathcal{V}}$, we recover the jiromaton and the minimal xor automaton, respectively. Finally, for $\mathcal{V}$ = distributive lattices we get a new canonical nfa called the *distromaton*, which bears a close resemblance to the universal automaton [20].

Example 1.1. Consider the language $L_n = (a+b)^*a(a+b)^n$ where $n \in \omega$. Its minimal dfa has 2^{n+1} states, see Example 2.12, and its state-minimal nfa has $n+2$ states. The átomaton, minimal xor automaton, jiromaton and distromaton of L_n are the nfas with at most $n+3$ states depicted below; see the Examples 3.20–3.23 for detailed explanations.



Átomaton — Minimal xor automaton

Jiromaton — Distromaton

Generally, the sizes of the four canonical nfas and the minimal dfa are related as follows:

(a) All the four canonical nfas can have exponentially fewer states than the minimal dfa.
(b) The minimal xor automaton and jiromaton have no more states than the minimal dfa.
(c) The átomaton and distromaton have the same number of states, although their structure can be very different. It can happen that the number of states is exponentially larger than that of the minimal dfa.