Information and Computation 239 (2014) 71–86

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Two-way automata making choices only at the endmarkers

Viliam Geffert [a,1], Bruno Guillon [b], Giovanni Pighizzini [c,*]

[a] Department of Computer Science, P.J. Šafárik University in Košice, Slovakia
[b] LIAFA, Université Paris Diderot, Paris 7, France
[c] Dipartimento di Informatica, Università degli Studi di Milano, Italy

ARTICLE INFO

Article history:
Received 24 July 2013
Available online 22 August 2014

Keywords:
Two-way automata
Descriptional complexity
Regular languages

ABSTRACT

The question of the state-size cost for simulation of two-way nondeterministic automata (2NFAS) by two-way deterministic automata (2DFAS) was raised in 1978 and, despite many attempts, it is still open. Subsequently, the problem was attacked by restricting the power of 2DFAS (e.g., using a restricted input head movement) to the degree for which it was already possible to derive some exponential gaps between the weaker model and the standard 2NFAS. Here we use an opposite approach, increasing the power of 2DFAS to the degree for which it is still possible to obtain a subexponential conversion from the stronger model to the standard 2DFAS. In particular, it turns out that subexponential conversion is possible for two-way automata that make nondeterministic choices only when the input head scans one of the input tape endmarkers. However, there is no restriction on the input head movement. This implies that an exponential gap between 2NFAS and 2DFAS can be obtained only for unrestricted 2NFAS using capabilities beyond the proposed new model.

As an additional bonus, conversion into a machine for the complement of the original language is polynomial in this model. The same holds for making such machines self-verifying, halting, or unambiguous. Finally, any superpolynomial lower bound for the simulation of such machines by standard 2DFAS would imply $L \neq NL$. In the same way, the alternating version of these machines is related to $L \stackrel{?}{=} NL \stackrel{?}{=} P$, the classical computational complexity problems.

1. Introduction

The cost, in terms of states, of the simulation of two-way nondeterministic automata (2NFAS, for short) by two-way deterministic automata (2DFAS) is one of the most important and challenging open problems in automata theory and, in general, in theoretical computer science. This problem was proposed in 1978 by Sakoda and Sipser [1], who conjectured that the cost is exponential. However, in spite of all effort, exponential gaps were proved only between 2NFAS and some restricted *weaker versions of* 2DFAS.

In 1980, Sipser proved that if the resulting machine is required to be *sweeping* (deterministic and reversing the direction of its input head only at the *endmarkers*, two special symbols used to mark the left and right ends of the input), the simulation of a 2NFA is indeed exponential [2]. However, Berman and Micali [3,4] proved independently that this does not solve

* Corresponding author.
E-mail addresses: viliam.geffert@upjs.sk (V. Geffert), guillonb@liafa.univ-paris-diderot.fr (B. Guillon), pighizzini@di.unimi.it (G. Pighizzini).

[1] Supported by the Slovak Grant Agency for Science under contract VEGA 1/0479/12 "Combinatorial Structures and Complexity of Algorithms", and by the Slovak Research and Development Agency under contract APVV-0035-10 "Algorithms, Automata, and Discrete Data Structures".

http://dx.doi.org/10.1016/j.ic.2014.08.009
0890-5401/

the general problem: in fact the simulation of unrestricted 2DFAs by sweeping 2DFAs also requires an exponential number of states. Sipser's result was generalized by Hromkovič and Schnitger [5], who considered *oblivious machines* (following the same trajectory of input head movements along all inputs of equal length) and, recently, by Kapoutsis [6], considering 2DFAs with the number of input head reversals that is sublinear in the length of the input. However, even the last condition gives a machine provably less succinct than unrestricted 2DFAs, and hence the general problem remains open.

Starting from 2003 with a paper by Geffert et al. [7], a different kind of restriction has been investigated in this context: the subclass or regular languages using *a single-letter input alphabet*. Even under this restriction, the problem of Sakoda and Sipser looks difficult, since it is connected with $L \stackrel{?}{=} NL$, an open question in complexity theory. (L and NL denote the respective classes of languages accepted in deterministic and nondeterministic logarithmic space.) First, in [7], a new normal form was obtained for unary automata, i.e., for automata with a single-letter input alphabet. In this form all nondeterministic choices and input head reversals take place only at the endmarkers. Moreover, the state-size cost of the conversion into this normal form is only *linear*. This normal form is a starting point for several other properties of unary 2NFAs. First, in the same paper, each n-state unary 2NFA is simulated by an equivalent 2DFA with $O(n^{\lceil \log_2(n+1)+3 \rceil})$ states, which gives a subexponential but still superpolynomial upper bound. It is not known whether this simulation is tight. However, a positive answer would imply the separation between the classes L and NL. In fact, under assumption that $L = NL$, each unary 2NFA with n states can be simulated by a 2DFA with a number of states polynomial in n [8]. After a minor modification (without assuming $L = NL$), this gives that each unary 2NFA can be made unambiguous, keeping the number of the states polynomial. (For further connections between two-way automata and logarithmic space, we address the reader to [9–11].)

Along these lines of investigation, in [12], the problem of the complementation for unary 2NFAs has been considered, by proving that each n-state 2NFA accepting a unary language L can be replaced by a 2NFA with $O(n^8)$ states accepting the complement of L. The proof combines the above normal form for unary 2NFAs with inductive counting arguments.

Kapoutsis [13] considered the complementation in the case of general input alphabets, but restricting the input head reversals. He showed that the complementation of *sweeping* 2NFAs (with the input head reversals only at the endmarkers) requires exponentially many states, thus emphasizing a relevant difference with the unary case.

In this paper, we use a different approach. Instead of *restricting the power of* 2DFAs to the degree for which it is already possible to derive an exponential gap between the weaker model and the standard 2NFAs, we *increase the power of* 2DFAs, towards 2NFAs, to the degree for which it is still possible to obtain a subexponential conversion from the stronger model to the standard 2DFAs. Such new stronger model then clearly shows that, in order to prove an exponential gap between 2NFAs and 2DFAs, one must use capabilities not allowed in the proposed new model. More precisely, in our new model, we neither restrict the cardinality of input alphabets, nor put any constraint on the input head movement, i.e., head reversals can take place at any input position. On the other hand, we permit nondeterministic choices only when the input head is scanning one of the endmarkers. We shall call such machine a *two-way outer-nondeterministic finite automaton* (2ONFA).

It turns out that this machine has its natural counterpart also in the case of *two-way alternating finite automata* (2AFAs), which is a *two-way outer-alternating finite automaton* (2OAFA), making both universal and existential choices only at the endmarkers. (For recent results on 2AFAs, see [14,15].)

We show that several results obtained for *unary* 2NFAs can be extended to 2ONFAs, and some of them even to the alternating version, 2OAFAs, with *any* input alphabet and *any* input head movement. In particular, we prove the following:

- Each n-state 2ONFA can be simulated by a *halting* two-way *self-verifying* finite automaton (2SVFA) [16] with $O(n^8)$ many states. (Self-verifying automata are nondeterministic automata with a restricted kind of nondeterminism—symmetric with respect to accepted languages and their complements. A more detailed description is given in Section 2.) This fact has two important implications:
  - The complementation of 2ONFAs can be done by using a polynomial number of states. Note the contrast with the above mentioned case of sweeping 2NFAs, studied in [13].
  - Each 2ONFA can be simulated by a *halting* 2ONFA using a polynomial number of states.
- Each n-state 2ONFA can be simulated by a 2DFA with $O(n^{\log_2(n)+7})$ states.
- If $L = NL$, then each n-state 2ONFA can be simulated by a 2DFA with a number of states polynomial in n. Hence, a superpolynomial lower bound for the simulation of 2ONFAs by 2DFAs would imply $L \neq NL$. (Unlike in [9], there are no restrictions on the length of potential witness inputs.)
- Each n-state 2ONFA can be simulated by an *unambiguous* 2ONFA with a polynomial number of states.
- If $L = P$, then each n-state 2OAFA can be simulated by a 2DFA with a number of states polynomial in n, with the same consequences as presented for $L \stackrel{?}{=} NL$. (P denotes, as usual, the class of languages recognizable by deterministic Turing machines in polynomial time.)
- Similarly, if $NL = P$, we get the corresponding polynomial conversion from 2OAFAs to 2NFAs.

These results are obtained by generalizing the constructions given originally for the unary case, in [7,12,8]. However, here we do not have a normal form that simplifies automata by restricting input head reversals to the endmarkers. Our generalization relies on a different tool, presented in the first part of the paper. Basically, we extend some techniques developed originally for deterministic devices [17,12] to machines with nondeterminism at the endmarkers. This permits us to check the existence of certain computation paths, including infinite loops, by the use of a linear number of states.