RESULTS ON THE AVERAGE STATE AND TRANSITION COMPLEXITY OF FINITE AUTOMATA ACCEPTING FINITE LANGUAGES

Hermann Gruber and Markus Holzer
Institut für Informatik, Technische Universität München,
Boltzmannstraße 3, D-85748 Garching bei München, Germany
e-mail: {gruberh,holzer}@in.tum.de

ABSTRACT

The study of descriptional complexity issues for finite automata dates back to the mid 1950's. One of the earliest results is that deterministic and nondeterministic finite automata are computationally equivalent, and that nondeterministic finite automata can offer exponential state savings compared to deterministic ones, see [11]—by the powerset construction one increases the number of states from n to 2^n, which is known to be a tight bound. Motivated by several applications and implementations of finite automata in software engineering, programming languages and other practical areas in computer science, the descriptional complexity of finite automata problems has gained new interest during the last decade. Tight upper bounds for the deterministic and nondeterministic state complexity of many operations on regular languages are known [8, 11, 12].

In many applications the regular languages are actually finite as, e.g., in natural language processing or constraint satisfaction problems in artificial intelligence. This prompted quite some research activity on finite languages—see [11] for an overview. Obviously, the length of the longest word in a finite language is a lower bound on the number of states of a finite automaton accepting a finite language. In fact it can be even exponential in the length of the longest word in the finite language as shown in [2, 5]. To be more precise, there is a finite language L over a binary alphabet whose longest word is of length n such that the minimal deterministic finite automaton accepting L needs $\Theta(\frac{2^n}{n})$ states. For the state savings for changing from a deterministic finite automaton to a nondeterministic finite automaton the bounds for automata accepting finite languages is slightly weaker than in the general case. In [10] it was shown that one can transform every nondeterministic finite automaton accepting a finite language into an equivalent deterministic finite automaton increasing the number of states from n to $\Theta(\sqrt{2^n})$, and this bound was shown to be sharp. More results on the state complexity of operations on finite languages can be found in [3, 8].

However, most of the work on descriptional complexity of regular languages yields worst-case results. To our knowledge, very few attempts have been made in order to understand certain aspects of the average behaviour of regular languages [1, 4, 6, 9]. Average-case complexity turns out to be much harder to determine than worst-case complexity, as it is currently unknown how many non-isomorphic automata of n states there are over a binary alphabet. For a recent survey on the problem of enumerating finite automata we refer to [7]. However, for finite automata with a singleton letter input alphabet the enumeration problem was solved in [9],

where also the average-case state complexity of operations on unary languages was studied. In this paper we concentrate on the average-case descriptional complexity of deterministic and nondeterministic finite automata accepting finite languages. By choosing a finite language L with given maximal word length uniformly at random, one can treat the size of the minimal deterministic or nondeterministic finite automaton accepting L as a random variable. Observe that our setup is different to that used in [9]. There deterministic finite automata are chosen at random among all n-state deterministic finite *automata*, whereas our setup is centered at *languages*. Due to this difference in the model, the results on finite languages cannot be directly compared to each other.

At first glance we show that almost all deterministic finite automata accepting finite languages over a binary input alphabet with word length at most n have state complexity $\Theta(\frac{2^n}{n})$, which is asymptotically like the worst-case. Then we introduce a stochastic process to generate finite languages, which is shown to be equivalent to the above mentioned setup choosing a finite language uniformly at random. This stochastic language generation process allows us to investigate operations on finite languages from the average-case point of view. It turns out that the expected value of the state complexity of a deterministic finite automaton accepting the union or intersection of two finite languages is larger than $\frac{4}{5} \cdot \frac{2^n}{n}$ on the average, as n tends to infinity. Similar bounds can be derived for the iteration of Boolean operations. Moreover, the average-case complexity of operations on finite unary languages is determined exactly. Finally, nondeterministic finite automata are considered. It turns out that the nondeterministic state complexity is in $\Theta(\sqrt{2^n})$ on the average, which is superior to the deterministic case with respect to the number of *states*. However, interestingly we show that the number of *transitions* needed is again $\Theta(\frac{2^n}{n})$ in most cases. Hence, the overall size, i.e., the length of a description of a finite state machine, is from the average-case complexity point of view the same for both deterministic and nondeterministic finite automata. Finally, we note that results similar to those for the binary case can be derived for larger alphabet sizes along the way outlined here.

References

[1] F. Bassino and C. Nicaud. Enumeration and random generation of accessible automata. Preprint, Institut d'électronique et d'informatique Gaspard-Monge, Unviesité de Marne-la-vallée, November 2005.

[2] C. Câmpeanu and W. H. Ho. The maximum state complexity for finite languages. *Journal of Automata, Languages and Combinatorics*, 9(2–3):189–202, September 2004.

[3] C. Câmpeanu, K. Culik II, K. Salomaa, and S. Yu. State complexity of basic operations on finite languages. In O. Boldt and H. Jürgensen, editors, *Proceedings of the 4th International Workshop on Implementing Automata*, number 2214 in LNCS, pages 60–70, Potsdam, Germany, July 1999. Springer.

[4] J.-M. Champarnaud and T. Paranthoën. Random generation of DFAs. *Theoretical Computer Science*, 330(2):221–235, 2005.

[5] J.-M. Champarnaud and J.-E. Pin. A maxmin problem on finite automata. *Discrete Applied Mathematics*, 23:91–96, 1989.

[6] M. Domaratzki. State compelxity of proportional removals. *Journal of Automata, Languages and Combinatorics*, 7(4):455–468, 2002.

[7] M. Domaratzki, D. Kisman, and J. Shallit. On the number of distinct languages accepted by finite automata with n states. *Journal of Automata, Languages and Combinatorics*, 7(4):469–486, 2002.

[8] M. Holzer and M. Kutrib. State complexity of basic operations on nondeterministic finite automata. In J.-M. Champarnaud and D. Maurel, editors, *Proceedings of the 7th International Conference Implementation and Application of Automata*, number 2608 in LNCS, pages 148–157, Tours, France, July 2003. Springer.

[9] C. Nicaud. Average state complexity of operations on unary automata. In M. Kutylowski, L. Pacholski, and T. Wierzbicki, editors, *Proceedings of the 24th International Symposium on Mathematical Foundations of Computer Science*, number 1672 in LNCS, pages 231–240, Szklarska Poreba, Poland, September 1999. Springer.

[10] K. Salomaa and S. Yu. NFA to DFA transformation for finite language over arbitrary alphabets. *Journal of Automata, Languages and Combinatorics*, 2(3):177–186, 1997.

[11] S. Yu. Regular languages. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 1, pages 41–110. Springer, 1997.

[12] S. Yu, Q. Zhuang, and K. Salomaa. The state complexity of some basic operations on regular languages. *Theoretical Computer Science*, 125:315–328, 1994.