Journal of Automata, Languages and Combinatorics 8 (2003) 2, 117–144

EFFICIENT ALGORITHMS FOR TESTING THE TWINS PROPERTY[1]

CYRIL ALLAUZEN and MEHRYAR MOHRI
AT&T Labs – Research
180 Park Avenue, Florham Park, NJ 07932, USA
e-mail: {allauzen, mohri}@research.att.com

ABSTRACT

Weighted automata and transducers are powerful devices used in many large-scale applications. The efficiency of these applications is substantially increased when the automata or transducers used are deterministic. There exists a general determinization algorithm for weighted automata and transducers that is an extension of the classical subset construction used in the case of unweighted finite automata [14]. However, not all finite-state transducers or weighted automata and transducers can be determinized using that algorithm, thus the question of the determinizability in that sense is essential. There exists a characterization of the determinizability of functional finite-state transducers and that of unambiguous weighted automata over the tropical semiring based on a general *twins property*. In the case of finite-state transducers, we give an efficient algorithm for testing functionality in time $O(|Q|^2\,|\Delta|+|E|^2)$ where Q is the set of states, E the set of transitions, and Δ the output alphabet of the input transducer. We also present a new and computationally more efficient algorithm for testing the twins property whose complexity is $O(|Q|^2(|Q|^2+|E|^2))$. In the automata case, we present a new and substantially more efficient algorithm for testing the twins property for unambiguous and cycle-unambiguous weighted automata over commutative and cancellative semirings whose complexity is $O(|Q|^2+|E|^2)$, which we conjecture to be optimal. Our experiments show our algorithms for testing the twins property to be practical with large weighted automata and transducers of several million transitions found in speech recognition applications.

Keywords: Finite-state transducers, weighted automata, rational power series, determinization, twins property, algorithms

1. Introduction

Finite automata are classical computational devices used in a variety of large-scale applications [1]. Some applications such as text, speech or image processing require more general devices, *weighted automata*, to account for the variability of the data and to rank alternative hypotheses [14, 9]. A weighted automaton is a finite automaton in which each transition carries some weight in addition to the usual symbol. Finite-state transducers are automata whose transitions are additionally labelled with an output label.

[1]Full version of a lecture presented at the Workshop *Weighted Automata: Theory and Applications* (Dresden University of Technology, Germany, March 4–8, 2002).