Journal of Automata, Languages and Combinatorics **22** (2017) 1–3, 5 – 28
© Institut für Informatik · Justus-Liebig-Universität Giessen

ON THE AVERAGE COMPLEXITY OF PARTIAL DERIVATIVE AUTOMATA FOR SEMI-EXTENDED EXPRESSIONS

Rafaela Bastos Sabine Broda António Machiavelo Nelma Moreira
Rogério Reis

CMUP & DCC, Faculdade de Ciências da Universidade do Porto
Rua do Campo Alegre 1024, 4250-007, Porto, Portugal
`{rrbastos,sbb}@dcc.fc.up.pt` `ajmachia@fc.up.pt` `{nam,rvr}@dcc.fc.up.pt`

ABSTRACT

Extended regular expressions (with complement and intersection) are used in many applications due to their succinctness. In particular, regular expressions extended with intersection only (also called semi-extended) can already be exponentially smaller than standard regular expressions or equivalent nondeterministic finite automata. For practical purposes it is important to study the average behaviour of conversions between these models. In this paper, we focus on the conversion of regular expressions with intersection to nondeterministic finite automata, using partial derivatives and the notion of support. We give a tight upper bound of $2^{O(n)}$ for the worst-case number of states of the resulting partial derivative automaton, where n is the size of the expression. Using the framework of analytic combinatorics, we establish an upper bound of $(1.056+o(1))^n$ for its asymptotic average-state complexity, which is significantly smaller than the one for the worst case. Some experimental results here presented suggest that, on average, the upper bound may not be exponential. Finally, we study the class of semi-extended regular expressions with only one occurrence of intersection at the top level. In this case, the worst-case state complexity of the partial derivative automaton is quadratic on the size of the expression, but we obtained an upper-bound that is, asymptotically and on average, $O(n^{\frac{3}{2}})$.

Keywords: regular expressions, intersection, automata, partial derivatives, average case complexity, analytic combinatorics

1. Introduction

Regular expressions with additional operators are used in applications such as programming languages [14], XML processing [28], or runtime verification [27]. Most of these operators do not increase their language expressive power but lead to gains

This work was partially supported by CMUP (UID/MAT/00144/2013), which is funded by FCT (Portugal) with national (MEC) and european structural funds through the programs FEDER, under the partnership agreement PT2020.