Journal of Automata, Languages and Combinatorics **22** (2017) 1–3, 169 – 188

STATE COMPLEXITY OF PREFIX DISTANCE OF SUBREGULAR LANGUAGES

Timothy Ng David Rappaport Kai Salomaa

School of Computing, Queen's University
Kingston, Ontario K7L 3N6, Canada
ng@cs.queensu.ca daver@cs.queensu.ca ksalomaa@cs.queensu.ca

ABSTRACT

The neighbourhood of a regular language of constant radius with respect to the prefix distance is always regular. We give upper bounds and matching lower bounds for the size of the minimal deterministic finite automaton (DFA) needed for the radius k prefix distance neighbourhood of an n state DFA that recognizes, respectively, a finite, a prefix-convex, a prefix-closed, a prefix-free, and a right ideal language. For prefix-closed languages the lower bound automata are defined over a binary alphabet. For finite and prefix-convex regular languages the lower bound constructions use an alphabet that depends on the size of the DFA and it is shown that the size of the alphabet is optimal.

Keywords: finite automata, state complexity, finite languages, prefix-convex languages, distance measures

1. Introduction

The neighbourhood of radius r of a language L consists of all words that are within distance at most r from some word of L. A distance measure d is said to be regularity preserving if the neighbourhood of any regular language with respect to d is regular. Calude et al. [3] have shown that *additive distances* are regularity preserving. Additivity requires, roughly speaking, that the distance is compatible with concatenation of words in a certain sense and best known examples of additive distances include the Levenshtein distance and the Hamming distance [3, 6].

The prefix distance of two words u and v is the sum of the lengths of the suffixes of u and v that begin after the longest common prefix of u and v. The suffix distance and the factor distance are defined analogously in terms of the longest common suffix (respectively, factor) of two words. It is known that the prefix, suffix and factor distance preserve regularity [5].

By the state complexity of a regularity preserving distance we mean the worst-case size of the minimal deterministic finite automaton (DFA) needed to recognize the radius r neighbourhood of an n state DFA language (as a function of n and r).