

TRANSDUCERS WITH SET OUTPUT¹

JUREK CZYZOWICZ

*Département d'informatique, Université du Québec en Outaouais
Gatineau, Québec J8X 3X7, Canada
e-mail: jurek@uqo.ca*

WOJCIECH FRACZAK

*IDT Canada, Ottawa, Ontario, Canada
e-mail: wojtek.fraczak@idt.com*

and

ANDRZEJ PELC

*Département d'informatique, Université du Québec en Outaouais
Gatineau, Québec J8X 3X7, Canada
e-mail: pelc@uqo.ca*

ABSTRACT

We consider transducers with set output, i. e., finite state machines which produce a set of output symbols (rather than a string of symbols) upon reading any input symbol. When a word consisting of input symbols is read, the union of corresponding output sets is produced. Such transducers are instrumental in some important data classification tasks, such as multi-field packet classification. Two transducers are called *equivalent* if they produce equal output upon reading any input word. In practical data classification applications, it is important to store in memory only one transducer of every equivalence class, in order to save memory space. This yields the need of finding, in any equivalence class, one transducer, called *canonical*, which is easy to compute, given any transducer from this class. The main result of this paper is the construction of an algorithm which completes this task. Assuming that the input and output alphabets are of bounded size, for a given n -state transducer T , our algorithm finds the canonical transducer $\Psi(T)$ equivalent to T in time $O(n \log n)$. The complexity of this algorithm is $O(n)$ if T is acyclic, i. e., when the input language for T is finite.

Keywords: Automaton, labelling, set output, transducer

1. Introduction

Classical (subsequential) transducers [5, 3, 7, 8] are finite automata which output strings of symbols after reading every input symbol, and concatenate these strings

¹A preliminary version of this paper appeared in the Proceedings of the 8th Annual International Conference on Computing and Combinatorics (COCOON 2002), Singapore, August 2002, LNCS 2387.