Deterministic Generalized Automata *

Dora Giammarresi** and Rosa Montalbano

Dipartimento di Matematica ed Applicazioni, Università di Palermo
via Archirafi, 34 - 90123 Palermo - ITALY
e-mail: {dora, rosalba}@altair.math.unipa.it

Abstract. A generalized automaton (GA) is a finite automaton where the single transitions are defined on words rather than on single letters. Generalized automata were considered by K. Hashiguchi who proved that the problem of calculating the size of a minimal GA is decidable. We define the model of deterministic generalized automaton (DGA) and study the problem of its minimization. A DGA has the restriction that, for each state, the sets of words corresponding to the transitions of that state are prefix sets. We solve the problem of calculating the number of states of a minimal DGA for a given language, by giving a procedure that effectively constructs it starting from the minimal (conventional) deterministic automaton.

1 Introduction

Generalized automata (GA) are a model of representation for regular languages that extends the notion of finite automata by allowing the single transitions to be defined on words rather than on single letters. Intuitively, a generalized automaton can be obtained from a conventional one by shrinking long paths of the graph in a unique edge with a "long" label. Therefore, generalized automata are usually more concise than conventional ones representing the same event.

In the past decades, several efforts have been devoted to compute the complexity of representation of a given language inside different models of representation (deterministic, non-deterministic, unambiguous, two-way, alternating, probabilistic, pebbles automata, regular expressions, logical formalisms and so on). The complexity of a language in a given model is generally understood as the size of the minimal representation of the language in that model. For example, a classical measure of the complexity of a finite automaton is its number of states and the complexity of a language in this model is the number of states of a minimal (with respect to the number of states) automaton recognizing it.

In this context, Hashiguchi in 1991 investigated the problem of computing the size of the minimal representation of a given regular language in the model.

* Work partially supported by the ESPRIT II Basic Research Actions Program of the EC under Project ASMICS 2 (contract No. 6317) and in part by MURST under project 40% Algoritmi, Modelli di Calcolo, Strutture Informative.

** Partially supported by a research fellowship by (INdAM).
of generalized automata (see [H91]). In particular, he proved that the problem of calculating the number of states of a minimal GA is actually decidable.

A strictly related problem consists of effectively computing a minimal representation of a given language in a model. In the case of conventional deterministic finite automata, it can be proved that the minimal automaton is unique and an algorithm to calculate it starting from any equivalent deterministic automaton can be obtained using the Myhill-Nerode's theorem (see, for example, [HU79]). For non-deterministic automata there are only partial results stating that there is no unique minimal automaton but there are no constructive procedures for computing it, excepting the one that lists all possible automata. In [JR91] the computational complexity of different problems concerning minimization is studied in a general setting for non-deterministic automata and it is proved that all these problems are computationally hard.

In this paper we introduce the model of deterministic generalized automata (DGA) and deal with the minimization problem for this model. In order to preserve all properties implied by the notion of determinism in the case of conventional automata, DGA have the restriction that the sets of words corresponding to the transitions of each state are prefix sets. We solve the problem of computing the number of states of a minimal DGA by giving a procedure to construct a minimal DGA for a given language starting from the minimal (conventional) deterministic one. We introduce two operations that allow one to reduce the number of states of a DGA: the first, called $I$-reduction, contracts states that are "indistinguishable" and the second, called $S$-reduction, suppresses states that are "superfluous". Then we give the conditions under which such operations can be performed. We show that there can be deterministic GA that are irreducible (with respect to the above operations) but not minimal and give necessary and sufficient conditions to reduce a deterministic GA to get a minimal one. Moreover, we show that, differently from the case of conventional deterministic automata, the minimal deterministic GA is not unique.

The size of the minimal representation of a language in a given model (which measures the complexity of the language) plays a primary role also in comparing different models according to their intrinsic succinctness. Much work has been devoted to studying succinctness of representation when transducers are considered (see, for example [WK94]). In the case of finite automata, very recently, Harel et al. studied exponential discrepancies in the succinctness of finite automata when augmented by combinations of various additional mechanisms like alternation (i.e. both universal and existential branching), concurrency, "two-wayness" and pebbles (see [GH94]). We conclude the paper by discussing problems of discrepancy in succinctness between non-deterministic and deterministic versions of generalized automata and give some open problems.

2 Preliminaries

We denote by $\Sigma$ a finite alphabet and by $\Sigma^*$ the free monoid generated by $\Sigma$. The elements of $\Sigma$ are called letters, those of $\Sigma^*$ are called words; the subsets