Review: Doctoral Thesis

Title: *Restricted Restarting Automata*

Author: Peter Černo

Reviewer: Martin Kutrib, Institut für Informatik, Universität Giessen, Giessen, Germany

June 28, 2015

The doctoral thesis under review is based on selected results by Peter Černo, obtained in automata theory and formal languages in connection with learnability from the perspective of context rewriting systems and restarting automata. In particular, it is based on his achievements on the expressive capacity and closure properties of clearing restarting automata and the grammatical inference of restricted context rewriting systems. A main focus of attention is set on the relations between the language families in question and the context-free languages, thus, covering naturally the range between the regular and growing context-sensitive languages.

The notion of restarting automata as it stands has a well known meaning in the field of (applications of) theoretical computer science. These devices enjoy a natural motivation as they have been introduced by Jančar, Mráz, Plátek, and Vogel (1995) in order to model the so-called "analysis by reduction," which is a technique used in linguistics to analyze sentences of natural languages that have a high degree of free word order. The technique consists of stepwise simplification of an extended sentence such that the (in)correctness of the sentence is not affected. In the past two decades since their introduction many aspects of variants of restarting automata have been subject to intensive research. Several well-known language families are characterized by some variant in a unified framework. Though a lot of results have been obtained, there are still some important open questions and a further development of the theory is natural. So, the investigation of restarting automata aspects is challenging on the one hand, and may extend our knowledge on basic concepts in computation theory on the other hand. In his thesis, Peter Černo presents new results (partially based on a selection from his single-authored and co-authored articles) he obtained on restricted versions of restarting automata in the general framework of context rewriting systems. He develops novel techniques that allow a concise and understandable description of technically involved constructions in connection with language acceptance. Approaches of several areas as learning theory, computational complexity, and string-rewriting systems are applied to achieve the goals of the thesis. The results obtained represent significant contributions to the development of the research area.

The thesis is organized in an introduction and six chapters. The short introduction provides the reader with the background of this line of research, a motivation, and an overview on further chapters. The first chapter introduces standard terminology of automata theory and formal languages, describes the relevant systems, and recapitulates basic results. In the second chapter, selected topics are covered that are closely related to the models under consideration. Since the automata considered are stateless and process their inputs mainly in a "free" order, they are considered to be string-rewriting systems. The relevant background of such systems is recalled and discussed. The last preliminary Chapter 3 (whose length is a little unbalanced) formulates the three main goals of the thesis. Having thus all notions carefully defined and tools provided, the next three chapters contain the main contributions of the thesis.

Chapter 4 deals with so-called clearing restarting automata, the most restricted variant considered. Essentially, these devices are string-rewriting systems that do not use auxiliary symbols and are only allowed to delete factors of the input iteratedly dependent on the context in which the factors appear. First it is shown that clearing restarting automata accept all regular languages. Then closure properties of the family of languages accepted are studied. It turned out that the family is not closed under several of the standard language operations. The non-closure properties are shown by (variants) of a context-free witness language that cannot by accepted by any clearing restarting automaton since a certain center marker has necessarily to be deleted in a reduction step. A further result is that the length of the contexts which may be used in the reduction rules to control the simplification process do matter. In particular, an infinite, tight, and proper content-length hierarchy is derived. The main part of this chapter is devoted to the comparison with context-free languages. The central question is for the necessary resources a clearing restarting automaton must have to accept non-context-free languages. In order to show that context-length four is sufficient, a novel representation technique is developed. The so-called circle-square representation of a language offers a simplified construction that is based on the, say, core structure of a language. Moreover, a scheme for learning instructions of a clearing restarting automaton from a given set of reductions is described. Both clever ideas are combined to obtain the desired result. The overall proof is technically quite involved and non-trivial. Further sections of the chapter complement this result as follows. It is shown that clearing restarting automata with context-length one cannot break the context-free barrier. However, there is a non-context-free language over a six-letter alphabet accepted with context-length two. The proof relies on another novel concept that allows to send signals through what is called ether and to recover the ether again. Finally, combining all these approaches, a clearing restarting automaton with context-length three is derived that accepts a non-context-free language over a two-letter alphabet. The chapter is concluded by a result concerning the computational complexity of the membership problem, it is shown to be $\mathcal{NP}$-complete, and a brief consideration of the extension to subword-clearing restarting automata.

Section 5 continues the investigation of context rewriting systems without auxiliary symbols. It is devoted to the grammatical inference of such systems. A general learning framework is developed that allows to infer effectively some types of λ-confluent models from positive and negative examples. Here a polynomial time of the learning paradigm is traded for a slight relaxation of the identification. Essentially, λ-confluence means that, if an input is to be accepted then every sequence of applications of simplification rules reduces the input to the empty word. For length-reducing rules this implies that the membership problem is solvable in linear time. The chapter provides a valuable introduction into the basics of learning theory. The algorithm is presented in detail. The ideas and procedures are complemented by several examples that improve the readability of the chapter significantly. First systems are considered whose context lengths as well as sizes of the simplification rules, that is, whose widths are restricted. Another main result in this chapter is the NP-hardness of the decision problem whether there exists a clearing restarting automaton with restricted width that is consistent with a given set of positive and negative examples. The proof is highly non-trivial.

In Chapter 6 context rewriting systems with auxiliary symbols are studied. Basically, only systems with one sole additional auxiliary symbol are considered. Depending on the number of this symbol on the right-hand side of any rewriting rule, Δ- and Δ^*-clearing restarting automata are distinguished. That is, in the former case at most one auxiliary symbol can be written and in the latter case the number of auxiliary symbols allowed is bounded by the length of the left-hand side. A first result in this chapter is that any context-free language is accepted by some Δ^*-clearing restarting automaton (with context length one). At first glance, the trick of the proof is immediate: simply encode every nonterminal symbol of a grammar in Chomsky normalform by a sequence of auxiliary symbols. A closer look at this idea reveals the problem that the construction has to be done with a finite number of instructions only. However, with careful preliminary considerations it is proved that this is always possible. In what follows a further highlight of the thesis is shown. The question is whether all context-free languages can be accepted by some Δ-clearing restarting automaton as well. That is, by a clearing restarting automaton that may write just one instance of the auxiliary symbol in every step. The question is answered in the affirmative by reduction of the construction for Δ^*-clearing restarting automata. In particular, the technically challenging proof uses an encoding of words where only symbols of the word may be rewritten by the sole auxiliary symbol. The development of the encoding is the key tool for showing the main result. It is again far from standard and highly non-trivial. The chapter is concluded by some results on a further generalization of clearing restarting automata to so-called limited context restarting automata that can use several auxiliary symbols.

Finally, in a concluding chapter Peter Černo summarizes the results obtained and discusses future directions for investigations.

The following questions could be asked during the defense of the thesis:

- Clearing restarting automata over unary alphabets characterize the unary regular languages. How about the sizes of the restarting automata compared with the sizes of equivalent deterministic or nondeterministic finite automata?

- Is the family of languages accepted by clearing restarting automata closed under complementation, Kleene star, or inverse homomorphism? If this is unknown, why is it hard to prove the closure properties?

- The membership problem for clearing restarting automata is decidable. How about further properties as, for example, emptiness, finiteness, equivalence, or inclusion?

- Do you know or expect a hierarchy of language families with respect to the width of the instructions?

• Become Δ-clearing restarting automata stronger when Δ is considred to be a finite set of auxiliary symbols?

Summarizing, the results presented in the thesis are important contributions to theoretical computer science, in particular, to automata theory, formal languages, and learning theory. Their significance lies both in the development and the use of new concepts and approaches as well as in the sharpening of results from the literature. The acceptance of the achievements demonstrated in the thesis by the international scientific community was positive. The articles the thesis is based on were published at scientific forums with high reputation. This emphasizes even more that the presented results are important and relevant not only for the direct community, but for the whole field of theoretical computer science. The presentation of the thesis is clear. The reader can easily follow the text, the statements are accompanied with many helpful comments, explanations, and examples. The writing quality is very good.

The thesis reflects and certifies at an extremely high level of focus technical skills and research talent of Peter Černo. He has written a remarkable doctoral thesis that presents exciting and thematically unified research on restarting automata and context rewriting systems. The goals stated in the thesis were reached and its results can be applied in numerous branches of science. Therefore, I strongly recommend to accept the thesis.

Giessen, June 28, 2015

M. Kutrib

Prof. Dr. Martin Kutrib