Journal of Automata, Languages and Combinatorics **28** (2023) 1–3, 143 – 166

TWO-DIMENSIONAL RANK-REDUCING GRAMMARS AND THEIR COMPLEXITY

Daniel Průša

Department of Cybernetics, Czech Technical University
Karlovo nám. 13, 121 35 Prague 2, Czech Republic
daniel.prusa@fel.cvut.cz

ABSTRACT

We study properties of a two-dimensional grammar introduced recently for use in document analysis and recognition. The grammar is obtained from the two-dimensional context-free grammar by limiting the form of productions. Variants (ranks) of the grammar with regard to productions complexity are defined. If the grammar is restricted to produce one-row pictures only, it generates regular languages. We suggest that the lowest-rank variant can be considered as a natural generalization of the regular matrix grammar, which in addition shares some of its good properties. Namely, it can be parsed in time linear in the input area and the emptiness problem is still decidable for the grammar. However, we also show that the higher-rank variants do not loosen complexity of the context-free grammar too much. There is a conditional lower bound preventing to propose a linear-time parsing algorithm. Moreover, the grammar is able to simulate the 2-counter Minsky machine, which results in non-recursive trade-offs between grammars of different ranks and also in undecidability of the emptiness problem.

Keywords: picture language, two-dimensional context-free grammar, regular matrix grammar, parsing complexity, decidability, descriptional complexity

1. Introduction

The two-dimensional (2D) context-free grammar, also known as Kolam grammar, was proposed by Siromoney et al. [24] and later independently by other authors [14, 22]. It is a natural generalization of the context-free grammar in the Chomsky normal form. It uses productions of the forms $N \to A$, $N \to A\,B$, $N \to \begin{smallmatrix} A \\ B \end{smallmatrix}$ to generate 2D arrays of symbols (so called *pictures*) by concatenating subpictures produced by A and B.

The grammar proved to be useful in the field of document analysis if it is relaxed to generate planar layouts of printed or handwritten characters. It was applied to recognition of domains such as mathematical formulas [19, 3], flowchart diagrams [12] or

The author was supported by the Czech Science Foundation grant no. 19-21198S.
0000-0003-4866-5709