Journal of Computer and System Sciences 81 (2015) 911–939

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Context-free coalgebras

Joost Winter [a,b], Marcello M. Bonsangue [a,b,*], Jan J.M.M. Rutten [a,c]

[a] Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, Netherlands
[b] LIACS, Niels Bohrweg 1, 2333 CA Leiden, Netherlands
[c] Radboud Universiteit Nijmegen, Heyendaalseweg 35, 6525 AJ Nijmegen, Netherlands

ARTICLE INFO

Article history:
Received 20 January 2013
Accepted 24 July 2013
Available online 3 December 2014

Keywords:
Context-free languages
Formal grammars
Bisimulation
Automatic sequence
Automata theory
Universal coalgebra

ABSTRACT

In this article, we provide a coalgebraic account of parts of the mathematical theory underlying context-free languages. We characterize context-free languages, and power series and streams generalizing or corresponding to the context-free languages, by means of systems of behavioural differential equations; and prove a number of results, some of which are new, and some of which are new proofs of existing theorems, using the techniques of bisimulation and bisimulation up to linear combinations. Furthermore, we establish a link between automatic sequences and these systems of equations, allowing us to, given an automaton generating an automatic sequence, easily construct a system of behavioural differential equations yielding this sequence as a context-free stream.

1. Introduction

During the last 15 years, a coalgebraic picture of the theory of automata and formal languages has emerged and been further developed. The earliest traces of this picture date back to work by Janusz A. Brzozowski who introduced the idea of *derivatives* of regular expressions in [5]. This idea of input derivatives was linked to the more abstract notion of an F-coalgebra for a functor F by the third author of the present article in [15], and further developed in e.g. [18,20,22].

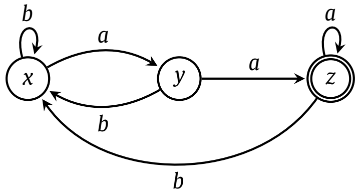
In [23], we extended this coalgebraic picture further, giving a characterization of the context-free languages as solutions to systems of *behavioural differential equations*, in which all Brzozowski derivatives are presented as polynomials over the set of variables, and in [4], we have generalized the theory to formal power series in noncommuting variables, providing a new characterization of so-called *algebraic* or *context-free* power series, coinciding with the familiar generalization from context-free languages to algebraic power series.

In the present article, we will provide a more systematic account of this framework. It turns out that important classes of formal languages and formal power series—finitary, rational, and algebraic (or context-free) series, respectively—can each be characterized as solutions to some format of systems of behavioural differential equations.

The simplest class of systems, which will be recalled in Section 2, consists of equations in which each derivative is presented simply as another state (or, equivalently, as a *variable* or *nonterminal*): these systems can be seen as describing deterministic automata and generalizations thereof. As an example of such a system, consider the deterministic automaton (presented without initial state)

* Corresponding author.
E-mail address: m.m.bonsangue@liacs.leidenuniv.nl (M.M. Bonsangue).

http://dx.doi.org/10.1016/j.jcss.2014.12.004

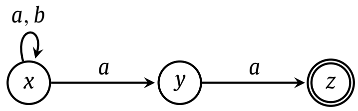


which corresponds to the following system of behavioural differential equations:

$$\begin{array}{lll} o(x) = 0 & x_a = y & x_b = x \\ o(y) = 0 & y_a = z & y_b = x \\ o(z) = 1 & z_a = z & z_b = x. \end{array}$$

Here, the equation $o(x) = 0$ can be read as 'x is a non-accepting state', $o(z) = 1$ as 'z is an accepting state', and $x_a = y$ as 'x makes an a-transition to y'.

In Section 3, we will combine the notions of automata and modules yielding the notion of a linear automaton, and recall some of the main ideas and results regarding rational power series and linear automata. This class of series can be characterized by systems of behavioural differential equations, in which every derivative is presented as a linear combination of states. For example, the nondeterministic automaton



is in direct correspondence with the system of behavioural differential equations

$$\begin{array}{lll} o(x) = 0 & x_a = x + y & x_b = x \\ o(y) = 0 & y_a = z & y_b = 0 \\ o(z) = 1 & z_a = 0 & z_b = 0. \end{array}$$

In this system, for example, the equation $x_a = x + y$ can be read as 'x makes a-transitions to x and y', and $y_b = 0$ can be read as 'y makes no b-transitions'.

In Section 4 we will consider a class of systems which yields a class of automata with an even richer structure than that of linear automata, and establish in Section 4.1 its equivalence with existing notions of *context-freeness* or *algebraicity*. Here, the derivatives of all nonterminals are presented as *polynomials* with coefficients in the underlying semiring.

An example of such a system is the following

$$\begin{array}{lll} o(x) = 1 & x_a = xy & x_b = 0 \\ o(y) = 0 & y_a = 0 & y_b = 1 \end{array}$$

and we will see in Section 4 that in this system x can be interpreted as the language $\{a^n b^n \mid n \in \mathbb{N}\}$. Although closely related to pushdown automata (with a single state), these systems correspond more directly to context-free grammars in Greibach normal form. For example, the system above is in direct correspondence with the following grammar:

$$\begin{array}{l} x \to 1 \mid axy \\ y \to b. \end{array}$$

Next, we will consider of a number of preservation results, and concrete constructions of new context-free systems of behavioural differential equations from old systems, providing some deeper links with the classical theory of context-free languages and automatic sequences. In Section 6, we will introduce the **zip** operation and a number of related operations, and show that these operations all preserve context-freeness. Finally, in Section 7, we will connect our notion of context-free power series to the theory of automatic sequences by showing that streams over finite fields are context-free if and only if they are algebraic, and furthermore present a simple method to extract systems of behavioural differential equations from automatic sequences.

The subject matter of this article can perhaps best be characterized as *coalgebraic automata theory*. Although closely related to the traditional, algebraic, approach to automata theory, and benefitting greatly from usage of traditional algebraic structures, our approach differs by making extensive usage of *bisimulations* (and bisimulations up to linear combinations), *systems of behavioural differential equations*, and *final coalgebra semantics*, and by altogether omitting the usage of matrices. Combined, these techniques enable us to provide proofs that are often more concise than the traditional ones, and different in style, often establishing close connections between the behaviour of automata, and algebraic properties characterizable by simple equations. Our approach also relates to the more abstract world of *universal coalgebra* [16], in the sense that many of