Information and Computation ••• (••••) •••–•••

Contents lists available at ScienceDirect

Information and Computation

www.elsevier.com/locate/yinco

Weighted automata and logics for infinite nested words

Manfred Droste *, Stefan Dück *,[1]

Institute of Computer Science, Leipzig University, D-04109 Leipzig, Germany

ARTICLE INFO

Article history:
Received 27 June 2014
Available online xxxx

Keywords:
Nested words
Weighted automata
Weighted logics
Quantitative automata
Valuation monoids

ABSTRACT

Nested words introduced by Alur and Madhusudan are used to capture structures with both linear and hierarchical order, e.g. XML documents, without losing valuable closure properties. Furthermore, Alur and Madhusudan introduced automata and equivalent logics for both finite and infinite nested words, thus extending Büchi's theorem to nested words. Recently, average and discounted computations of weights in quantitative systems found much interest. Here, we will introduce and investigate weighted automata models and weighted MSO logics for infinite nested words. As weight structures we consider valuation monoids which incorporate average and discounted computations of weights as well as the classical semirings. We show that under suitable assumptions, two resp. three fragments of our weighted logics can be transformed into each other. Moreover, we show that the logic fragments have the same expressive power as weighted nested word automata.

1. Introduction

Nested words, introduced by Alur and Madhusudan [2], capture models with both a natural sequence of positions and a hierarchical nesting of these positions. Prominent examples include XML documents and executions of recursively structured programs. Automata on nested words, logical specifications, and corresponding languages of nested words have been intensively studied, see [1,2,22]. Recently, there has been much interest in quantitative features for the specification and analysis of systems. Quantitative automata modeling the long-time average or discounted behavior of systems were investigated by Chatterjee, Doyen, and Henzinger [6,7]. It is the goal of this paper to present quantitative logics for such quantitative automata on nested words.

The connection between MSO logic and automata due to Büchi, Elgot, and Trakhenbrot [5,20,28] has proven most fruitful. Weighted automata over semirings (like $(\mathbb{N}, +, \cdot, 0, 1)$) were already investigated by Schützenberger [26] and soon developed a flourishing theory, cf. the books [3,19,21,25] and the recent handbook [12]. However, an expressively equivalent weighted MSO logic was developed only recently [10]. This was extended to semiring-weighted automata and logics over finite nested words in [23], and further to strong bimonoids as weight structures in [14]. For quantitative automata and logics, incorporating average and discounting computations of weights over words, such an equivalence was given in [13].

In this paper, we will investigate quantitative nested word automata and suitable quantitative MSO logics. We will concentrate on infinite nested words, although our results also hold for finite nested words. We employ the stair Muller nested word automata of [2,22] since these can be determinized without losing expressive power. As weight structures, we take the valuation monoids of [13]. These include infinite products as in totally complete semirings [15], but also

* Corresponding authors.
E-mail addresses: droste@informatik.uni-leipzig.de (M. Droste), dueck@informatik.uni-leipzig.de (S. Dück).

[1] Supported by Deutsche Forschungsgemeinschaft (DFG), project DR 202/11-1 and Graduiertenkolleg 1763 (QuantLA).

http://dx.doi.org/10.1016/j.ic.2016.06.010
0890-5401/

Please cite this article in press as: M. Droste, S. Dück, Weighted automata and logics for infinite nested words, Inf. Comput. (2016), http://dx.doi.org/10.1016/j.ic.2016.06.010

computations of long-time averages or discountings of weights. As example for such a setting, we give the calculation of the long-time ratio of bracket-free positions in prefixes of an infinite nested word. As our first main result, we show that under suitable assumptions on the valuation monoid D, two resp. three versions of our weighted MSO logic have the same expressive power. In particular, if D is commutative, then any weighted MSO-formula is equivalent to one in which conjunctions occur only between 'classical' boolean formulas and constants. In contrast to [13], our proof uses direct conversions of the formulas and thus has much better complexity than using the automata-theoretic constructions of [13]. These conversions are new even for the case of weighted logics on words.

In our second main result, we show under suitable assumptions on the valuation monoid that our weighted MSO logics have the same expressive power as weighted nested word automata. These assumptions on the valuation monoid are satisfied by long-time average resp. discounted computations of weights; therefore our results apply to these settings. All our constructions of automata from formulas and conversely are effective.

An extended abstract of this paper appeared in [9].

2. Automata and logics for nested ω-words

In this section, we describe basic background for classical (unweighted) automata and logics on nested ω-words. We denote by Σ a finite alphabet and by Σ^ω the set of all ω-words over Σ. We denote with $\mathbb{N}$ the set of all natural numbers without zero. For a binary relation R, we denote with $R(x, y)$ that $(x, y) \in R$.

Definition 1. A *matching relation* ν over $\mathbb{N}$ is a subset of $(\{-\infty\} \cup \mathbb{N}) \times (\mathbb{N} \cup \{\infty\})$ such that:

(i) $\nu(i, j) \Rightarrow i < j$,
(ii) $\forall i \in \mathbb{N} : |\{j : \nu(i, j)\}| \leq 1 \wedge \ |\{j : \nu(j, i)\}| \leq 1$,
(iii) $\nu(i, j) \wedge \nu(i', j') \wedge i < i' \Rightarrow j < i' \vee j > j'$,
(iv) $(-\infty, \infty) \notin \nu$.

A *nested ω-word nw* over Σ is a pair $(w, \nu) = (a_1 a_2 ..., \nu)$ where $w = a_1 a_2 ...$ is an ω-word over Σ and ν is a matching relation over $\mathbb{N}$. We denote by $NW^\omega(\Sigma)$ the set of all nested ω-words over Σ, and we call every subset of $NW^\omega(\Sigma)$ a *language of nested ω-words*.

If $\nu(i, j)$ holds, we call i a *call position* and j a *return position*. In case of $j = \infty$, i is a *pending call* otherwise a *matched call*. In case of $i = -\infty$, j is a *pending return* otherwise a *matched return*. Note that similar to [2], condition (iii) ensures that nestings of calls and returns are either disjoint or hierarchical; in particular, a position cannot be both a call and a return. If i is neither call nor return, then we say i is an *internal*.

Definition 2. A *deterministic stair Muller nested word automaton (sMNWA)* over Σ is a quadruple $\mathcal{A} = (Q, q_0, \delta, \mathfrak{F})$, where $\delta = (\delta_{\text{call}}, \delta_{\text{int}}, \delta_{\text{ret}})$, consisting of:

- a finite set of states Q,
- an initial state $q_0 \in Q$,
- a set $\mathfrak{F} \subseteq 2^Q$ of accepting sets of states,
- the transition functions $\delta_{\text{call}}, \delta_{\text{int}} : Q \times \Sigma \to Q$,
- the transition function $\delta_{\text{ret}} : Q \times Q \times \Sigma \to Q$.

A *run* r of the sMNWA $\mathcal{A}$ on the nested ω-word $nw = (a_1 a_2 ..., \nu)$ is an infinite sequence of states $r = (q_0, q_1, ...)$ where $q_i \in Q$ for each $i \in \mathbb{N}$ and q_0 is the initial state of $\mathcal{A}$ such that for each $i \in \mathbb{N}$ the following holds:

$$\begin{cases} \delta_{\text{call}}(q_{i-1}, a_i) = q_i & \text{, if } \nu(i, j) \text{ for some } j > i \\ \delta_{\text{int}}(q_{i-1}, a_i) = q_i & \text{, if } i \text{ is an internal} \\ \delta_{\text{ret}}(q_{i-1}, q_{j-1}, a_i) = q_i & \text{, if } \nu(j, i) \text{ for some } 1 \leq j < i \\ \delta_{\text{ret}}(q_{i-1}, q_0, a_i) = q_i & \text{, if } \nu(-\infty, i) \ . \end{cases}$$

We call $i \in \mathbb{N}$ a *top-level position* if there exist no positions $j, k \in \mathbb{N}$ with $j < i < k$ and $\nu(j, k)$. We define

$$Q^t_\infty(r) = \{q \in Q \mid q = q_i \text{ for infinitely many top-level positions } i\} \ .$$

A run r of an sMNWA is *accepted* if $Q^t_\infty(r) \in \mathfrak{F}$. An sMNWA $\mathcal{A}$ *accepts* the nested ω-word nw if there is an accepted run of $\mathcal{A}$ on nw. We denote with $L(\mathcal{A})$ the set of all accepted nested ω-words of $\mathcal{A}$. We call a language L of nested ω-words *regular* if there is an sMNWA $\mathcal{A}$ with $L(\mathcal{A}) = L$.

Alur and Madhusudan [2] considered nondeterministic Büchi NWA and nondeterministic Muller NWA. They showed that the deterministic versions of these automata have strictly less expressive power than the nondeterministic automata. However, referring to Löding, Madhusudan, and Serre [22], Alur and Madhusudan stated that deterministic stair Muller NWA

Please cite this article in press as: M. Droste, S. Dück, Weighted automata and logics for infinite nested words, Inf. Comput. (2016), http://dx.doi.org/10.1016/j.ic.2016.06.010