Information Processing Letters 111 (2011) 500–502

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Recognizability of the support of recognizable series over the semiring of the integers is undecidable

D. Kirsten [a], K. Quaas [b,*]

[a] Institut für Informatik, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany
[b] Institut für Informatik, Universität Leipzig, 04009 Leipzig, Germany

ARTICLE INFO

Article history:
Received 14 December 2010
Received in revised form 4 February 2011
Accepted 25 February 2011
Available online 1 March 2011
Communicated by A. Muscholl

Keywords:
Formal languages
Formal power series
Weighted automata
Post's correspondence problem
Supports

ABSTRACT

A recognizable series over the semiring of the integers is a function that maps each word over an alphabet to an integer. The support of such a series consists of all those words which are not mapped to zero. It is long known that there are recognizable series whose support is not a recognizable, but a context-free language. However, the problem of deciding whether the support of a given recognizable series is recognizable was never considered. Here we show that this problem is undecidable. The proof relies on an encoding of an instance of Post's correspondence problem.

1. Introduction

A *formal power series* (series, for short) over a semiring is a function from the set of all words over an alphabet to the semiring. A series S over a semiring is called *recognizable* if there is a weighted finite automaton over the semiring whose behavior corresponds to S. One stream in the rich theory of series deals with the connection to formal languages. The *support* of a series is the set of all words that are not mapped to zero. Supports of recognizable series, or equivalently, weighted finite automata, have been extensively studied (see e.g. [2,1,6]). In particular, one is interested in generalizations of classical decision problems like language equivalence, emptiness and universality. For instance, Eilenberg [2] shows that given two recognizable series S_1 and S_2 over the semiring of the rationals, it is decidable whether S_1 equals S_2. In the proof of this result he uses that it is decidable whether the support of a given recognizable series over the semiring of the rationals is empty [2]. In contrast to this, it is not decidable whether the support of a given recognizable series over the rationals equals the set of all finite words [1]. However, it is also of great interest to decide, given a recognizable series S, whether the support of S is recognizable by a finite automaton. It is long known (see e.g. [7]) that for a recognizable series S over so-called positive semirings (e.g. the min–plus-semiring), the support of S is *always* recognizable. The same is true for recognizable series over locally finite semirings [1,6] and commutative, quasi-positive semirings [4]. On the other hand, there are semirings for which there are recognizable series whose support is *not* recognizable. This is for instance the case for the semiring of the integers. With this in mind, two questions arise. First, can one find a characterization of classes of semirings for which the support of a recognizable series is always recognizable (called *SR-semirings* in the following)? Second, given a recognizable series S over a non-SR-semiring (e.g. the semiring of the integers), can one decide whether the support of S is recognizable? Considering the first question, one of the authors recently gave an algebraic characterization for SR-semirings [3]. The goal of this paper is to move towards the answer of the second question: We prove that, given a recognizable series S over

* Corresponding author.
E-mail address: quaas@informatik.uni-leipzig.de (K. Quaas).

0020-0190/$ – see front matter
doi:10.1016/j.ipl.2011.02.014

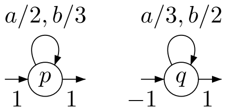

Fig. 1. A weighted finite automaton whose support is not recognizable.

the semiring of the integers, it is not decidable whether the support of S is recognizable.

2. Preliminaries and main result

Let $\mathbb{Z}$ denote the set of integers. We consider the semiring $(\mathbb{Z}, +, \cdot, 0, 1)$ over the integers with usual addition and multiplication and unit elements 0 and 1, respectively.

Let Σ be some finite alphabet. We denote the empty word by ε. We use $|w|$ to denote the length of a word $w \in \Sigma^*$ and $|w|_a$ to denote the number of a occurring in w. By Σ^+ we mean the set $\Sigma^* \backslash \{\varepsilon\}$.

A *series* is a function $S : \Sigma^* \to \mathbb{Z}$. A *weighted finite automaton* over $\mathbb{Z}$ is a tuple $(Q, T, \mu, \lambda, \varrho)$, where

- Q is a non-empty, finite set of states,
- $T \subseteq Q \times \Sigma \times Q$ is a set of transitions,
- $\mu : T \to \mathbb{Z}$ is a function assigning to each transition a weight,
- $\lambda, \varrho : Q \to \mathbb{Z}$ is a function assigning to each state an initial respectively an accepting weight.

Let $w = a_1 \dots a_n$ be a finite word over Σ. A *run* of $\mathcal{A}$ on w is a finite sequence $(q_0, a_1, q_1)(q_1, a_2, q_2) \dots (q_{n-1}, a_n, q_n)$ of transitions such that $(q_{i-1}, a_i, q_i) \in T$ for each $i \in \{1, \dots, n\}$. We use $p \overset{w}{\rightsquigarrow} q$ to denote the set of runs on w with $q_0 = p$ and $q_n = q$. Let $\bar{\mu} : T^* \to \mathbb{Z}$ be the unique extension of μ to a homomorphism. Each run r is assigned the weight $\bar{\mu}(r)$ in $\mathbb{Z}$. The *behavior* of $\mathcal{A}$ on w is the series $\|\mathcal{A}\| : \Sigma^* \to \mathbb{Z}$ defined by $\|\mathcal{A}\|(w) := \sum_{p,q \in Q,\, r \in p \overset{w}{\rightsquigarrow} q} \lambda(p) \cdot \bar{\mu}(r) \cdot \varrho(q)$. We say that a series S is *recognizable* if there is a weighted finite automaton $\mathcal{A}$ over $\mathbb{Z}$ such that $\|\mathcal{A}\| = S$.

For $L \subseteq \Sigma^*$, we define the *characteristic series* $1_L : \Sigma^* \to \mathbb{Z}$ by $1_L(w) = 1$ if $w \in L$ and $1_L(w) = 0$ otherwise. We recall that for each recognizable language $L \subseteq \Sigma^*$, the characteristic series 1_L is recognizable: If L is recognizable, then there is a deterministic automaton $\mathcal{A}$ such that $L(\mathcal{A}) = L$. We obtain a weighted finite automaton $\mathcal{A}'$ from $\mathcal{A}$ by assigning weight 1 to each transition of $\mathcal{A}$ and setting $\lambda(p) = 1$ if p is an initial state in $\mathcal{A}$, and $\lambda(p) = 0$ otherwise. Similarly, we set $\varrho(p) = 1$ if p is a final state, and $\varrho(p) = 0$ otherwise. We clearly have $\|\mathcal{A}'\| = 1_L$. We also recall that recognizable series are closed under sum [2].

We define the support of a series $S : \Sigma^* \to \mathbb{Z}$ as $\mathrm{supp}(S) = \{w \in \Sigma^* \mid S(w) \neq 0\}$. The following example is well known.

Example 1. Let $S : \Sigma^* \to \mathbb{Z}$ be the series defined by

$$S(w) = 2^{|w|_a} 3^{|w|_b} - 3^{|w|_a} 2^{|w|_b}.$$

This series is recognizable, see Fig. 1. For each $w \in \Sigma^*$, we have $S(w) = 0$ if and only if $|w|_a = |w|_b$. Hence, $\mathrm{supp}(S) = \{w \in \Sigma^* \mid |w|_a \neq |w|_b\}$.

Given a finite word $w = a_1 \dots a_n \in \{0, 1\}^*$, we use $\mathrm{val}(w)$ to denote the integer $\sum_{1 \leqslant i \leqslant n} a_i \cdot 2^{n-i}$, i.e., the integer that is binarily represented by w. We let $\mathrm{num} : \{0, 1\}^* \to \mathbb{Z}$ be the mapping defined by

$$\mathrm{num}(w) = 2^{|w|} + \mathrm{val}(w)$$

for each $w \in \{0, 1\}^*$. Intuitively, num maps each word w over $\{0, 1\}$ to the integer that is binarily represented by the word $1w$, i.e., the integer denoted by $\mathrm{val}(1w)$. Notice that num is injective.

For each $z \in \mathbb{Z}$, we use $\mathrm{bit}_i(z)$ to denote the i-th least significant bit in the binary representation of the *absolute* value of z.

Let A be a finite alphabet and let $\alpha, \beta : A^* \to \{0, 1\}^*$ be two homomorphisms. We say that $w \in A^+$ is a *solution* of the triple (A, α, β) if $\alpha(w) = \beta(w)$. *Post's correspondence problem* [5] (PCP) asks, given the triple (A, α, β), does (A, α, β) have a solution? This problem is undecidable [5].

In this paper, we consider the following problem.

SUPPORT RECOGNIZABILITY PROBLEM
Input: A recognizable series $S : \Sigma^* \to \mathbb{Z}$
Question: Is $\mathrm{supp}(S)$ recognizable by a finite automaton?

Theorem 1. *The support recognizability problem is undecidable.*

The proof of Theorem 1 is a reduction of PCP: given an instance (A, α, β) of PCP, we define a recognizable series S such that (A, α, β) has no solution if and only if $\mathrm{supp}(S)$ is recognizable by a finite automaton (Lemma 4). The details are given in the next section.

3. Proof of the main result

Let (A, α, β) be an instance of PCP, where A is a finite alphabet and $\alpha, \beta : A^* \to \{0, 1\}^*$ are homomorphisms. We further let $k \geqslant 1$ be such that for each $a \in A$ we have $k > |\alpha(a)| + 1$ and $k > |\beta(a)| + 1$.

Let $\Sigma = A \,\dot\cup\, \{\#, b, c\}$, where $\#, b, c$ are symbols not occurring in A. We define the series $S : \Sigma^* \to \mathbb{Z}$ as follows: Let $u \in A^+$, $n_b, n_c \geqslant 1$. Then

$$S\left(u\#b^{n_b}c^{n_c}\right) = \mathrm{num}\left(\alpha(u)\right) - \mathrm{num}\left(\beta(u)\right) + \left(2^{2k}\right)^{|u|+n_b} - \left(2^{2k}\right)^{|u|+n_c}.$$

For each $w \notin A^+\#b^+c^+$ we let $S(w) = 1$.

Lemma 2. *The series S is recognizable.*

Proof. Note that $S = S_1 + S_2 + S_3 + S_4 + S_5$, where

$$S_1(w) = \begin{cases} \mathrm{num}(\alpha(u)) & \text{if } u \in A^+\#b^+c^+, \\ 0 & \text{otherwise,} \end{cases}$$

$$S_3(w) = \begin{cases} (2^{2k})^{|u|+n_b} & \text{if } u \in A^+\#b^{n_b}c^+, \\ 0 & \text{otherwise,} \end{cases}$$

for some $n_b \geqslant 0$, and

$$S_5(w) = \begin{cases} 1 & \text{if } u \notin A^+\#b^+c^+, \\ 0 & \text{otherwise,} \end{cases}$$

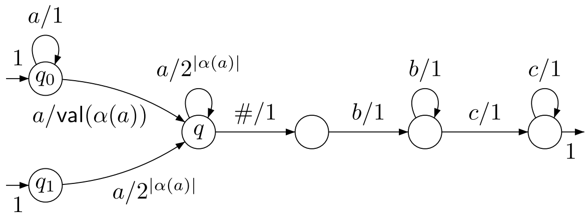

Fig. 2. $\mathcal{A}_1$. An edge label $a/\mathsf{val}(\alpha(a))$, e.g., indicates that for every $a \in A$, there is a transition with label a and weight $\mathsf{val}(\alpha(a))$.

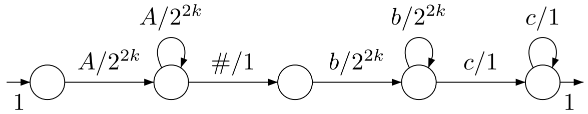

Fig. 3. $\mathcal{A}_3$.

and S_2 and S_4 are defined analogously. In the following we show that each of these series is recognizable. The lemma follows from the fact that recognizable series are closed under sum [2].

Let $\mathcal{A}_1$ be the weighted finite automaton depicted in Fig. 2. Notice that for each $w = u\#b^{n_b}c^{n_c}$ with $u \in A^+$ and $n_b, n_c \geqslant 1$, we have $\|\mathcal{A}_1\|(w) = \mathsf{num}(\alpha(u))$: Let $u = a_1 \ldots a_n$. Then there is a run $q_0 \xrightarrow{a_1} q_0 \cdots \xrightarrow{a_{i-1}} q_0 \xrightarrow{a_i} q \xrightarrow{a_{i+1}} q \cdots \xrightarrow{a_n} q$ for each $1 \leqslant i \leqslant n$. The weight of this run is obviously $\mathsf{val}(\alpha(a_i)) \cdot 2^{|\alpha(a_{i+1} \ldots a_n)|}$. Hence the sum over all n such runs equals $\mathsf{val}(\alpha(u))$. Addition with $2^{|u|}$ is modeled by the run starting in q_1. We further observe that $\|\mathcal{A}_1\|(w) = 0$ for every $w \in \Sigma^* \backslash A^+\#b^+c^+$. Hence, S_1 is recognizable. A weighted finite automaton recognizing S_2 can be constructed analogously.

We let $\mathcal{A}_3$ be the weighted finite automaton shown in Fig. 3. One can easily see that $\|\mathcal{A}_3\|(w) = (2^{2k})^{|u|+n_b}$ if $w = u\#b^{n_b}c^{n_c}$ for some $u \in A^+, n_b, n_c \geqslant 1$, and $\|\mathcal{A}_3\|(w) = 0$ otherwise. Hence, S_3 is recognizable. A weighted finite automaton recognizing S_4 can be defined analogously. Finally, we notice that S_5 is the characteristic series of a recognizable language and thus recognizable. $\square$

Lemma 3. *For every $u\#b^{n_b}c^{n_c}$ with $u \in A^+$ and $n_b, n_c \geqslant 1$, we have*

$$S\big(u\#b^{n_b}c^{n_c}\big) = 0 \quad \textit{iff} \quad \alpha(u) = \beta(u) \textit{ and } n_b = n_c.$$

Proof. The direction from the right to the left is obvious. We present the proof for the other direction. Let $w = u\#b^{n_b}c^{n_c}$ for some $u \in A^+$ and $n_b, n_c \geqslant 1$ and assume $S(w) = 0$. Then we have $\mathsf{bit}_i(\mathsf{num}(\alpha(u)) - \mathsf{num}(\beta(u))) = 0$ for each $i > 2 \cdot k \cdot |u|$ and $\mathsf{bit}_i((2^{2k})^{|u|+n_b} - (2^{2k})^{|u|+n_c}) = 0$ for each $i < 2 \cdot k \cdot |u|$. Hence $\mathsf{num}(\alpha(u)) = \mathsf{num}(\beta(u))$ and $n_b = b_c$. By injectivity of num, we further obtain $\alpha(u) = \beta(u)$. $\square$

Lemma 4. *The following assertions are equivalent*:

1. (A, α, β) *has no solution.*
2. $\mathsf{supp}(S) = \Sigma^*$.
3. $\mathsf{supp}(S)$ *is recognizable.*

Proof. "1. $\Rightarrow$ 2." Let $w = u\#b^{n_b}c^{n_c}$ for some $u \in A^+$ and $n_b, n_c \geqslant 1$. Since (A, α, β) has no solution, we have $\alpha(u) \neq \beta(u)$. By Lemma 3, $S(w) \neq 0$ and thus $w \in \mathsf{supp}(S)$. By definition of S, for each $w' \in \Sigma^* \backslash A^+\#b^+c^+$, we have $S(w') = 1$ and thus $w' \in \mathsf{supp}(S)$. Hence, $\mathsf{supp}(S) = \Sigma^*$.

"3. $\Rightarrow$ 1." By contradiction, let $u \in A^+$ such that $\alpha(u) = \beta(u)$. By Lemma 3, we observe $\mathsf{supp}(S) \cap u\#b^+c^+ = \{u\#b^{n_b}c^{n_c} \mid n_b,\ n_c \geqslant 1,\ n_b \neq n_c\}$, which is not recognizable by a straightforward pumping argument. Hence, $\mathsf{supp}(S)$ is not recognizable. $\square$

4. Open questions

Our proof generalizes obviously to all semirings which include $\mathbb{Z}$ as a subsemiring. However, the decidability of the support recognizability problem remains open for non-SR-semirings which do not include $\mathbb{Z}$. We do not even know whether such a semiring exists. Further, by the best of our knowledge the complexity of the decidable [2] problem to decide whether the support of a recognizable series over the semiring of the integers is empty is unknown.

References

[1] J. Berstel, C. Reutenauer, Noncommutative Rational Series With Applications, Encyclopedia of Mathematics and Its Applications, Cambridge University Press, 2010.

[2] S. Eilenberg, Automata, Languages and Machines, vol. A, Academic Press, 1974.

[3] D. Kirsten, An algebraic characterization of semirings for which the support of every recognizable series is recognizable, in: R. Královic, D. Niwinski (Eds.), MFCS, in: LNCS, vol. 5734, Springer, 2009, pp. 489–500.

[4] D. Kirsten, The support of a recognizable series over a zero-sum free commutative semiring is recognizable, in: V. Diekert, D. Nowotka (Eds.), DLT 2009, Proceedings, in: LNCS, vol. 5583, Springer, 2009, pp. 326–333.

[5] E. Post, A variant of a recursively unsolvable problem, Bull. Amer. Math. Soc. 52 (4) (1946) 264–268.

[6] J. Sakarovitch, Rational and recognisable power series, in: M. Droste, W. Kuich, H. Vogler (Eds.), Handbook of Weighted Automata, in: EATCS Monographs in Theoretical Computer Science, Springer, 2009, pp. 105–174.

[7] A. Salomaa, M. Soittola, Automata-Theoretic Aspects of Formal Power Series, Springer, New York, 1978.