Theoretical Computer Science 36 (1985) 113–117
North-Holland

NOTE

ON EQUALITY OF MULTIPLICITY SETS OF REGULAR LANGUAGES

Keijo RUOHONEN

Institute of Mathematics, Tampere University of Technology, SF-33101, Tampere 10, Finland

Communicated by M. Nivat
Received May 1983
Revised June 1984

Abstract. The equivalence problem for multiplicity sets of regular languages is shown to be undecidable.

1. Introduction

A *set with multiplicity* is a set K equipped with a universal set $U \supset K$ and a *multiplicity function* $M: U \to \mathbb{N}$ ($\mathbb{N}$ denotes the positive semiring of nonnegative integers). The usual interpretation is that an element α of U belongs to K with multiplicity αM (or "αM times"). The *multiplicity set* of a set K with multiplicity is UM.

As far as regular languages over a finite alphabet Σ are concerned, there is a natural way of defining multiplicity: U equals Σ^* and αM equals the number of ways in which α is accepted by a (nondeterministic) finite automaton recognizing the language. This is the approach taken in [3, Chapter VI]. (For another approach, see [9, Theorem II.5.6].) The class of such multiplicity functions equals the class of functions $\Sigma^* \to \mathbb{N}$ having *matrix representations*, that is, functions for which there exist a positive integer k, a vector $I \in \mathbb{N}^k$ (the *initial vector*), a morphism μ from Σ^* to the monoid of endomorphisms (*matrices*) on $\mathbb{N}^k$ and a morphism $T: \mathbb{N}^k \to \mathbb{N}$ (the *terminal morphism*) such that

$$\alpha M = I\alpha\mu T \quad \text{for all } \alpha \in \Sigma^*.$$

We show that it is undecidable of two bounded regular languages with multiplicities whether or not their multiplicity sets are equal.

2. The result

We begin with a lemma on representation of polynomials.

0304-3975/85/$3.30

Lemma 2.1. *Let $p(n_1, \ldots, n_m)$ be a polynomial with coefficients in $\mathbb{N}$. Then there exist (effectively) a positive integer k, a vector $I \in \mathbb{N}^k$, endomorphisms $E_1, \ldots, E_m$ on $\mathbb{N}^k$, and a morphism $T: \mathbb{N}^k \to \mathbb{N}$ such that*

$$p(n_1, \ldots, n_m) = IE_1^{n_1} \ldots E_m^{n_m} T.$$

Proof. Cf. [9, Theorem II.11.1]. □

Theorem 2.2. *It is undecidable of two bounded regular languages with multiplicities whether or not their multiplicity sets are equal.*

Proof. Let $p_1(n_1, \ldots, n_m)$ and $p_2(n_1, \ldots, n_m)$ be a pair of (arbitrary) polynomials with coefficients in $\mathbb{N}$. The unsolvability of Hilbert's tenth problem implies that there is no algorithm that decides of each such pair the solvability of the equation

$$p_1(n_1, \ldots, n_m) = p_2(n_1, \ldots, n_m) \tag{1}$$

in $\mathbb{N}$. It suffices, therefore, to construct two bounded regular languages with multiplicities such that their multiplicity sets are equal if and only if (1) has no solutions in $\mathbb{N}$.

Let $S, A_1, \ldots, A_k, B_1, \ldots, B_k, C, D, F$ be a set of generators of $\mathbb{N}^{2k+4}$. By Lemma 2.1 we can give representations of $1+p_1$ and $1+p_2$ over $\langle A_1, \ldots, A_k \rangle$ and $\langle B_1, \ldots, B_k \rangle$, respectively, for a sufficiently large positive integer k, say

$$1 + p_1(n_1, \ldots, n_m) = I_1 E_{11}^{n_1} \ldots E_{1m}^{n_m} T_1$$

and

$$1 + p_2(n_1, \ldots, n_m) = I_2 E_{21}^{n_1} \ldots E_{2m}^{n_m} T_2$$

(here $\langle V \rangle$ denotes the submonoid of $\mathbb{N}^{2k+4}$ generated by $V \subset \mathbb{N}^{2k+4}$).

We then define the endomorphisms $J_1, J_2, J_3, G_1, \ldots, G_m, H_1, H_2$ on $\mathbb{N}^{2k+4}$ as follows. We give only parts of the definitions, the remaining parts being arbitrary.

$$\begin{array}{lll}
SJ_1 = I_1 + C + I_2, & & \\
SJ_2 = D + C, & DJ_2 = D, & CJ_2 = D + C, \\
SJ_3 = C + F, & FJ_3 = F, & CJ_3 = C + F, \\
CG_i = C, & A_jG_i = A_jE_{1i}, & B_jG_i = B_jE_{2i}, \\
DH_1 = D, & FH_1 = F, & CH_1 = D + C + F, \\
CH_2 = C, & A_jH_2 = A_jT_1D, & B_jH_2 = B_jT_2F.
\end{array}$$

We then have

$$(1 + p_1(n_1, \ldots, n_m))D + C + (1 + p_2(n_1, \ldots, n_m))F = SJ_1 G_1^{n_1} \ldots G_m^{n_m} H_2$$

and, for every pair $p, q \in \mathbb{N}$ such that $p > q > 0$, we have

$$pD + C + qF = SJ_2^{p-q}H_1^q \quad \text{and} \quad qD + C + pF = SJ_3^{p-q}H_1^q.$$

Consider then $\mathbb{N}^{(2k+4)^2}$ as the tensor square of $\mathbb{N}^{2k+4}$ embedded in $\mathbb{N}^{2k+4} \times \mathbb{N}^{(2k+4)^2} = \mathbb{N}^{(2k+4)(2k+5)}$. Define the endomorphism H_3 on $\mathbb{N}^{(2k+4)(2k+5)}$ as follows:

$$\begin{aligned} D^{(2)}H_3 = F^{(2)}H_3 &= (D \otimes F)H_3 = (F \otimes D)H_3 = (C \otimes D)H_3 = (D \otimes C)H_3 \\ &= FH_3 = DH_3 = S, \\ C^{(2)}H_3 &= (C \otimes F)H_3 = (F \otimes C)H_3 = CH_3 = 0 \end{aligned}$$

(again, only the essential parts of the definition are given; $\otimes$ denotes tensor product and $X^{(2)}$ denotes the tensor square of X). We have

$$\begin{aligned} &(S \times S^{(2)})(J_1 \times J_1^{(2)})(G_1 \times G_1^{(2)})^{n_1} \dots (G_m \times G_m^{(2)})^{n_m}(H_2 \times H_2^{(2)})H_3 = \\ &\quad = r(1 + p_1(n_1, \dots, n_m), 1 + p_2(n_1, \dots, n_m))S \end{aligned}$$

and

$$\begin{aligned} (S \times S^{(2)})(J_2 \times J_2^{(2)})^{p-q}(H_1 \times H_1^{(2)})^q H_3 &= r(p, q)S, \\ (S \times S^{(2)})(J_3 \times J_3^{(2)})^{p-q}(H_1 \times H_1^{(2)})^q H_3 &= r(q, p)S, \end{aligned}$$

where $r(x, y) = (x + y)^2 + 3x + y$. It is well known that r is injective when restricted on $\mathbb{N}^2$.

Let then $\{j_1, j_2, j_3, g_1, \dots, g_m, h_1, h_2, h_3\}$ be an alphabet. Since

$$R_1 = j_2^+ h_1^+ h_3 + j_3^+ h_1^+ h_3 \quad \text{and} \quad R_2 = R_1 + j_1 g_1^* \dots g_m^* h_2 h_3$$

are bounded regular languages, we know, by the above (and by [9, Theorems II.4.5 and II.5.1]) that

$$L_1 = \{0\} \cup \{r(p, q) \mid p, q \in \mathbb{N}, p > 0, q > 0 \text{ and } p \neq q\}$$

and

$$L_2 = L_1 \cup \{r(1 + p_1(n_1, \dots, n_m), 1 + p_2(n_1, \dots, n_m)) \mid n_1, \dots, n_m \in \mathbb{N}\}$$

are multiplicity sets of bounded regular languages with multiplicities. Since r is injective on $\mathbb{N}^2$, we have the equality $L_1 = L_2$ if and only if (1) does not have a solution in $\mathbb{N}$. $\square$

3. Further discussion

In accordance with the definition of a commutative $\mathbb{N}$-rational series in [5] we define a *commutative language with multiplicity* as a language with multiplicity such that

$$\alpha \text{ is obtained from } \beta \text{ by permuting symbols } \Rightarrow$$

$$\Rightarrow \alpha \text{ and } \beta \text{ have the same multiplicity.}$$

An analogy of a theorem of Ginsburg and Spanier [4, Theorem 2.1] then holds true: A language over $\{a_1, \ldots, a_n\}$ with multiplicity is a commutative regular language with multiplicity if and only if it is the commutative closure of a regular subset of $a_1^* \ldots a_n^*$ with multiplicity (cf. [9, Theorem II.11.1 and its proof]). Hence, we have, by Theorem 2.2 and its proof, the following corollary.

Corollary 3.1. *It is undecidable of two commutative regular languages with multiplicities whether or not their multiplicity sets are equal.*

The cardinality of the alphabet can be reduced to two because we have the following lemma.

Lemma 3.2. *Any multiplicity set of a (bounded) regular language with multiplicity is also the multiplicity set of a (bounded) binary regular language with multiplicity.*

Proof. Let $\alpha M = I\alpha\mu T$ be a matrix representation of the multiplicity function M of a regular language K with multiplicity over $\{a_1, \ldots, a_n\}$. Let the representation be over $\mathbb{N}^k = \langle A_1, \ldots, A_k \rangle$.

Let then A_{ij} ($i = 1, \ldots, n$ and $j = 1, \ldots, k$) be a set of generators of $\mathbb{N}^{nk}$ and consider $\mathbb{N}^k$ as embedded in $\mathbb{N}^{nk}$ by identifying A_j with A_{1j}. Define then the endomorphisms E_1 and E_2 on $\mathbb{N}^{nk}$ by

$$A_{ij}E_1 = A_j a_i \mu,$$

$$A_{ij}E_2 = A_{i+1,j} \quad \text{for } i < n,$$

$$A_{nj}E_2 = A_{1j},$$

and the morphism $T_1 : \mathbb{N}^{nk} \to \mathbb{N}$ by $A_{ij}T_1 = A_j T$.

It is easily verified that the multiplicity set of a binary regular language K_1 with multiplicity represented by I, E_1, E_2 and T_1 equals $\{a_1, \ldots, a_n\}^* M$. It is also easily seen that if K is bounded, K_1 may be assumed to be bounded, too. □

Corollary 3.3. *It is undecidable of two bounded binary regular languages with multiplicities whether or not their multiplicity sets are equal.*

To the author's knowledge, the case is open for unary regular languages with multiplicities. For a rather large subclass of multiplicity sets of unary regular languages with multiplicities, the equality problem is, however, known to be decidable. This is the class of the so-called D0L growth sets, for which the equality problem was shown to be decidable by Berstel and Nielsen [1]. Extension of D0L growth sets to larger alphabets gives the so-called DT0L growth sets. (For D0L growth functions and DT0L growth functions, see [9, Sections III.7–8].) This gives some interest to the following corollary which can be proved by slightly modifying the proof of Theorem 2.2.

Corollary 3.4. *The equality problem for DT0L growth sets is undecidable.*

Proof. We first note that DT0L growth functions are closed under multiplication and that the characteristic function of the set of prefixes of $R_1 + j_1 g_1^* \dots g_m^* h_2$ (respectively R_2) is a DT0L growth function. It then suffices to take a positive terminal morphism T defined by $CT = 1$ and $AT = 2$ for all other generators A of $\mathbb{N}^{(2k+4)(2k+5)}$ (cf. [9, Exercise III.7.6]). The idea is that whenever $t > 0$ belongs to the multiplicity set, $2t$ is in the DT0L growth set, and that, apart from 0 and 4, these are the only even numbers in the growth set. □

Finally, we want to mention that the proof of Theorem 2.2 is largely inspired by the well-known proofs of the fact that the equality problem for sentential forms of linear grammars is undecidable, independently discovered by Blattner [2], Rozenberg [6], and Salomaa [8]. Rozenberg [7] used an analogous construction to prove that the equality problem for the so-called DT0L languages is undecidable. Our construction can be used to give a new proof to this fact. Only slight modifications in the proofs of Theorem 2.2 and Corollary 3.4 are needed. (These are, in fact, simplifications since no tensor squaring is needed and H_3 as well as the terminal morphism can be omitted.) Due to the 'commutative nature' of our proof, analogous modifications will show that the so-called *Parikh equivalence problem for DT0L languages is undecidable*, too; a result which appears to be new.

References

[1] J. Berstel and M. Nielsen, The growth range equivalence problem for D0L systems is decidable, in: A. Lindenmayer and G. Rozenberg, eds., *Automata, Languages, Development* (North-Holland, Amsterdam, 1976) 161–178.

[2] M. Blattner, The unsolvability of the equality problem for the sentential forms of context-free grammars, *J. Comput. System Sci.* 7 (1973) 463–468.

[3] S. Eilenberg, *Automata, Languages and Machines, Vol. A* (Academic Press, New York, 1974).

[4] S. Ginsburg and E.H. Spanier, Bounded regular sets, *Proc. AMS* 17 (1966) 1043–1049.

[5] J. Karhumäki, Remarks on commutative N-rational series, *Theoret. Comput. Sci.* 5 (1977) 211–217.

[6] G. Rozenberg, Direct proofs of the undecidability of the equivalence problem for sentential forms of linear context-free grammars and the equivalence problem for OL systems, *Inform. Process. Lett.* 1 (1972) 233–235.

[7] G. Rozenberg, The equivalence problem for deterministic TOL systems is undecidable, *Inform. Process. Lett.* 1 (1972) 201–204; Errata, id., 252.

[8] A. Salomaa, On sentential forms of context-free grammars, *Acta Inform.* 2 (1973) 40–49.

[9] A. Salomaa and M. Soittola, *Automata-Theoretic Properties of Formal Power Series* (Springer, Berlin, 1978).