Cybernetics and Systems Analysis, Vol. 40, No. 6, 2004

## BRIEF COMMUNICATIONS

# LANGUAGE RECOGNITION BY TWO-WAY DETERMINISTIC PUSHDOWN AUTOMATA

**L. P. Lisovik and D. A. Koval'**[†]

UDC 519.716.35

*It is proved that any bounded context-free language can be recognized by a two-way deterministic automaton with a finite-rotary counter.*

**Keywords:** *semilinear set, automaton with a finite-rotary counter, context-free language.*

The theory of formal languages has been developed intensively for more than 40 years. However, several intriguing questions are yet to be answered. Can a two-way deterministic pushdown automaton recognize a context-free (CF) language *L*, for one (see [1–3])? We will demonstrate here that a two-way deterministic automaton with a finite-rotary counter can recognize any bounded CF-language $L$. This result can be generalized to any bounded language whose Parique image is a semilinear set (Theorem 4). Let us first prove the following theorem.

**THEOREM 1.**Let $L \subseteq w_1^* \cdot \ldots \cdot w_n^*$ be an arbitrary bounded CF-language. Then it can be recognized by an appropriate two-way deterministic pushdown automaton $A$.

**Proof.**We will restrict ourselves to the case $L \subseteq a_1^* \cdot \ldots \cdot a_n^*$, where $a_i$, $1 \le i \le n$, are different symbols. Any bounded CF-language is known to be characterized by a finite union of linear sets with stratified systems of periods and, hence, by a finite union of linear sets with linearly independent stratified systems of periods (see [4]). Moreover, the class of languages recognizable by two-way deterministic pushdown automata is closed with respect to the operation of union. Thus, it will suffice to consider the case of a bounded CF-language $L$ whose characteristic is a linear set $L(\theta, P)$, where $\theta$ is a zero vector, and $P$ is a linearly independent stratified system of periods.

Any period $q \in P$ has no more than two nonzero coordinates. Let $L(\theta, P) \subseteq N^n$. Let us consider a set of numbers $\{1, \ldots, n\}$. The numbers $1 \le i \le j \le n$ are said to be interconnected if there exists a sequence $i_1, \ldots, i_k$, where $i = i_1$, $j = i_k$, and for $1 \le t < k$ the system $P$ contains a vector $q_t$ with the nonzero $i_t$th and $i_{t+1}$th coordinates. The case where $k = 2$, $i = j = i_1 = i_2$ is possible. It can be seen that the problem can be reduced to the case where the connectedness condition is satisfied: any two numbers $1 \le i < j \le n$ are interconnected.

A sequence of numbers $i_1, \ldots, i_k$ is called a ring if the above condition is satisfied, with $i_1 = i_k$, and the numbers $i_2, \ldots, i_{k-1}$ are pairwise different. Since the system $P$ is linearly independent, there is no more than one ring (accurate to torsion). Therefore, the problem is reduced to the case where the system $P$ is linearly independent, the connectedness condition holds, and no more than one ring exists.

We will call the numbers $1, \ldots, n$ nodes, the components (nodes) that belong to the ring the ring nodes, and the nodes with one outgoing edge the end nodes. The other nodes are called intermediate. All of this pertains to the system $P$ that satisfies the connectedness condition, i.e., the graph $\Gamma(P)$, which consists of the nodes $1, \ldots, n$ and the edges that correspond to the periods of the system *P*, has a path from any node *i* to any node *j*. Moreover, the graph $\Gamma(P)$ has either no ring paths or only one ring. It should be pointed out that an edge (loop) from a node into the same node is a ring. Thus, a logical analysis leads us to a finite number of graphs that actually constitute what may be called a pattern. Hereafter, it will suffice to consider only one graph $\Gamma(P)$ (however, all possible alternatives will be studied).

---

T. G. Shevchenko National University, Kiev, Ukraine, [†]*dkoval@aval.ua.* Translated from Kibernetika i Sistemnyi Analiz, No. 6, pp. 177-181, November-December 2004. Original article submitted October 10, 2004.

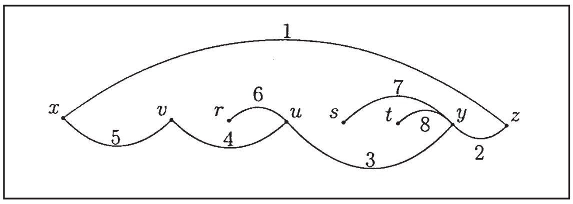


Fig. 1

It remains to consider the case where the system $P$ is linearly independent, the connectedness condition holds, and no more than one ring exists. Moreover, $P$ is a stratified system of periods, $P \subseteq N^n$. Let $i_1, \ldots, i_k$ be a ring defined by the system $P$. Since it is possible to renumber the nodes, the condition $1 \le i_1 < i_2 < \ldots < i_{k-1} \le n$ can be considered to hold. Recall that $i_1 = i_k$ by the definition of a ring. We will restrict the consideration to the case where $n = 8$ and the graph $\Gamma(P)$ has the form shown in Fig. 1.

Then the task is to use a word $a_1^{k_1} \cdot \ldots \cdot a_8^{k_8}$ to check whether $(k_1, \ldots, k_8) \in L(\theta, P)$, where the system $P$ consists of eight vectors, $P = \{p_i \mid 1 \le i \le 8\}$. For the sake of definiteness, let $p_i = (p_{i1}, \ldots, p_{i8})$, the vector $p_i$ having a pair of nonzero coordinates $(\delta_{i1}, \delta_{i2})$, where $(\delta_{i1}, \delta_{i2})$ is $(1, 8)$, $(7, 8)$, $(4, 7)$, $(2, 4)$, $(1, 2)$, $(3, 4)$, $(5, 7)$, and $(6, 7)$, respectively, for $1 \le i \le 8$. Let the numbers $1, \ldots, 8$ correspond to the variables $x$, $v$, $r$, $u$, $s$, $t$, $y$, and $z$. Let $c_{ij} = p_{i\delta_{ij}}$, $1 \le i \le 8$, $j = 1, 2$.

It is necessary to check whether there are the numbers $x_1$, $x_2$, $v_1$, $v_2$, $r$, $u_1$, $u_2$, $u_3$, $s$, $t$, $y_1$, $y_2$, $y_3$, $y_4$, $z_1$, and $z_2$ such that $k_1 = x_1 + x_2$, $k_2 = v_1 + v_2$, $k_3 = r$, $k_4 = u_1 + u_2 + u_3$, $k_5 = s$, $k_6 = t$, $k_7 = y_1 + y_2 + y_3 + y_4$, and $k_8 = z_1 + z_2$. Moreover, the following conditions must be satisfied:

$$\frac{x_1}{c_{11}} = \frac{z_2}{c_{12}}, \quad \frac{z_1}{c_{22}} = \frac{y_4}{c_{21}}, \quad \frac{y_3}{c_{82}} = \frac{k_6}{c_{81}},$$

$$\frac{y_2}{c_{72}} = \frac{k_5}{c_{71}}, \quad \frac{u_3}{c_{31}} = \frac{y_1}{c_{32}}, \quad \frac{u_2}{c_{62}} = \frac{k_3}{c_{61}},$$

$$\frac{u_1}{c_{42}} = \frac{v_2}{c_{41}}, \quad \frac{v_1}{c_{52}} = \frac{x_2}{c_{51}}.$$

Let us use the above scheme of reasoning, which has yielded the equation $x_2 = T(x_2)$ in solving the problem $a^{k_1} b^{k_2} c^{k_3} \in L_{13}$. Proceeding in a similar manner, we arrive at the equation $x_2 = T(x_2)$, where

$$T(x_2) = \frac{c_{51}}{c_{52}} \left( k_2 - \frac{c_{41}}{c_{42}} \left( k_4 - \frac{c_{62}}{c_{61}} k_3 - T_1(x_2) \right) \right),$$

$$T_1(x_2) = \frac{c_{31}}{c_{32}} \left( k_7 - \frac{c_{72}}{c_{71}} k_5 - \frac{c_{82}}{c_{81}} k_6 - \frac{c_{21}}{c_{22}} \left( k_8 - \frac{c_{12}}{c_{11}} (k_1 - x_2) \right) \right).$$

After transformations, the equation $x_2 = T(x_2)$ takes the form $x_2 = \beta$, where $\beta = \sum_{i=1}^{8} d_i' k_i$, where $d_i'$ are rational numbers.

Thus $(k_1, \ldots, k_8) \in L(\theta, P)$ if and only if $\beta$ is an integer non-negative number satisfying the condition $0 \le \beta \le k_1$. Whether these properties of $\beta$ are valid can be checked using a two-way deterministic counter automaton.

Thus, Theorem 1 is proved. From the proof, we can derive stronger statements.

**THEOREM 2.** Let $L$ be an arbitrary bounded CF-language. Then it can be recognized by a two-way deterministic automaton with a finite-rotary counter.

**Proof.** Let $L$ be a bounded CF-language, $L \subseteq w_1^* \cdot \ldots \cdot w_n^*$, where $w_i$, $1 \le i \le n$, are some words in an alphabet $\Sigma$. The proof of Theorem 1 involved the special case where $w_i = a_i$, $1 \le i \le n$, are different symbols. Let us now consider the general case.

Let $M = \{(k_1, \ldots, k_n) \mid w_1^{k_1} \cdot \ldots \cdot w_n^{k_n} \in L,\ k_i \in N,\ 1 \le i \le n\}$. Then, according to [4], $M$ is a semilinear set with stratified systems of periods. Using the corresponding theorems from [4], we obtain $M = \bigcup_{i=1}^{m} L_i$, where $L_i = L(c_i, P_i)$, $1 \le i \le m$, are linear sets with linearly independent stratified systems of periods. The class of languages recognized by two-way deterministic automata with a finite-rotary counter is closed with respect to the operation of finite union. Therefore, the problem reduces to the case $M = L_1$, where $L_1 = (\theta, P)$, and $\theta$ is a zero vector and $P$ is a linearly independent stratified system of periods.

In the general case, two different vectors $(k_1, \ldots, k_n)$, $(L_1, \ldots, l_n)$, such that $w = w_1^{k_1} \cdot \ldots \cdot w_n^{k_n}$, $w = w_1^{l_1} \cdot \ldots \cdot w_n^{l_n}$, can exist for the same word $w \in L$. In other words, the word $w$ is divided into $n$ blocks, $w = u_1, \ldots, u_n$, where $u_i = w_i^{k_i}$, $1 \le i \le n$. However, the way such a division can be done is not unique. For example, when the words $w_1$ and $w_2$ commutate with each other.

To keep the proof of Theorem 1, we need conventions that would provide unambiguous definition of the coefficients $k_1, \ldots, k_n$. Finally, the condition $w \in L$ will be equivalent to the condition $w \in \bigcup_{i=1}^{t} M_i$, where $M_i = L \cap v_{i1}^{*} \cdot \ldots \cdot v_{is_i}^{*}$, $1 \le i \le t$. Here, the vector $(k_1, \ldots, k_{s_i})$, where $w = v_{i1}^{k_1} \cdot \ldots \cdot v_{is_i}^{k_{s_i}}$ is defined uniquely for any $i$, $1 \le i \le t$, for any word $w$, $w \in v_{i1}^{*} \cdot \ldots \cdot v_{is_i}^{*}$.

The numbers $t$ and $s_i$, $1 \le i \le t$, and the system of words $v_{ij}$, $1 \le j \le s_i$, $1 \le i \le t$, are specified as follows. All these objects do not depend on the language $L$ and are defined by the system of words $w_1, \ldots, w_n$ alone.

This can be explained as follows. Let there exist a vector $(l_1, \ldots, l_n)$ such that $w = w_1^{l_1} \cdot \ldots w_n^{l_n}$. Then we obtain a situation (one of a finite number of situations) $i$, $1 \le i \le t$, such that $w = v_{i1}^{k_1} \cdot \ldots \cdot v_{is_i}^{k_{s_i}}$. One of the sequences $v_{i1}, \ldots, v_{is_i}$ occurs as follows.

Let $v_{i1} = w_1$. Assume that $w = w_1$ and $k_1$ is the maximum possible number such that $w = w_{i1}^{k_1} v_{i2}^{k_2} \cdot \ldots \cdot v_{is_i}^{k_{s_i}}$. Only the following cases are possible here:

(i) $s_i = 1$;

(ii) $s_i = 2$, $|v_{i2}| \le |w_1|, k_2 = 1$;

(iii) $s_i \ge 3$, $|v_{i2}| = |w_1|$, $v_{i2} \ne w_1$.

The case (ii) splits into subcases depending on the value of $v_{i2}$. The case (iii) also splits into subcases depending on the value of $v_{i2}$ and on the further strategy. The strategy is determined by the expected position of the subword $v_{i1}^{k_1}$ as a prefix of the word $w_1^{l_1} \cdot \ldots \cdot w_n^{l_n}$, i.e., any of the conditions of the form $\sum_{i=1}^{r} l_i |w_i| \le k_1 |v_{i1}| < \sum_{i=1}^{r+1} l_i |w_i|$ is assumed, and if $l_1 \ne 0$, then $l_1 |w_1| \le k_1 |v_{i1}|$. Thus, we obtain $w = w_1^{l_1} \cdot \ldots \cdot w_r^{l_r} uv$, where $u = u_1 u_2$, $v_{i1}^{k_1} = w_1^{l_1} \cdot \ldots \cdot w_r^{l_r} u_1$, $u_2 = u' w_{r+1}^{c} w_{r+2}^{l_{r+2}} \cdot \ldots \cdot w_n^{l_n}$, $|u'| < |w_{r+1}|$, $u'$ is an ending of the word $w_{r+1}$. After the objects $v_{i2}$, $u'$, and $r$ are fixed, it is assumed that the number $k_2$ is maximum possible, etc. It can be shown that all possible cases are exhausted in a finite number of such sequential refinements of the strategy.

This completely defines the number $t$, the numbers $s_i$, $1 \le i \le t$, and the systems of words $v_{ij}$, $1 \le j \le s_i$, $1 \le i \le t$.

Now the algorithm of checking the condition $w \in L$ suggests sequential checking of the conditions $w \in M_i$, where $M_i = L \cap v_{i1}^{*} \cdot \ldots \cdot v_{is_i}^{*}$, $1 \le i \le t$. We have $w \in L$ if at least one of the conditions $w \in M_i$, $1 \le i \le t$ holds. For any number $i$, $1 \le i \le t$, for a word $w \in M_i$ there exist only one vector $(k_1, \ldots, k_{s_i})$ such that $w = v_{i1}^{k_1} \cdot \ldots \cdot v_{is_i}^{k_{s_i}}$. Therefore, the condition $w \in M_i$ is checked by means of a two-way deterministic automaton with a one-rotary counter, following the scheme of proving Theorem 1. This completes the proof of Theorem 2.

**THEOREM 3.** Let $L$ be a finite union of finite intersections of bounded CF-languages. Then the language $L$ can be recognized by a two-way deterministic automaton with a finite-rotary.

**THEOREM 4.** Let $L \subseteq w_1^* \cdot \ldots \cdot w_n^*$ be a bounded language, where $\{(k_1, \ldots, k_n) \mid w_1^{k_1} \cdot \ldots \cdot w_n^{k_n} \in L,\ k_i \geq 0,$ $1 \leq i \leq n\}$ is a semilinear set. Then it can be recognized by a two-way deterministic automaton with a finite-rotary counter.

**Proof.**Any semilinear set can be represented as a finite intersection of semilinear sets with stratified systems of periods [5]. Therefore, the language $L$ is a finite intersection of bounded CF-languages, and Theorem 3 can be used.

If the bounded language $L$ is recognized by a two-way deterministic automaton with a finite-rotary stack, then it also can be recognized by a two-way deterministic automaton with a finite-rotary counter (see [6]). Problems of emptiness and equivalence are solvable in the class of languages recognized by a two-way deterministic automata with a finite-rotary counter [7].

There are many nontrivial examples of languages, including CF-languages, that can be recognized by two-way deterministic pushdown automata (see [8]). In conclusion, consider the following theorem without proof.

**THEOREM 5**. Let $n$ be a natural number. Let $L = \{x \mid x = v_1 \ldots v_n z$, where $v_i$ $(1 \leq i \leq n)$ are palindromes, $v_i \in \{0,1\}^+$, $z \in \{0,1\}^*\}$. Then each of the languages $L_n$ can be recognized by a two-way deterministic pushdown automaton.

We assume that there exists a CF-language that cannot be recognized by a two-way deterministic automaton, and we can give appropriate counterexamples, which, however, require additional analysis.

## REFERENCES


1. Z. Galil, "Some open problems in the theory of computation as questions about two-way deterministic pushdown automaton languages," Math. Syst. Theory, **10**, 211–228 (1977).
2. W. Rytter, "A hardest language recognized by two-way nondeterministic pushdown automata," Inform. Proc. Lett., 13, 145-146 (1981).
3. P. Duris and Z. Galil, "Fooling a two-way automaton or one pushdown store is better than one counter for two-way machines," Theoret. Comput. Sci., **21**, 39–53 (1982).
4. S. Ginzburg, Mathematical Theory of Context-Free Languages [Russian translation], Mir, Moscow (1970).
5. L. Y. Liu and P. Weiner, "A characterization of semilinear sets," J. Comput. System. Sci., No. 4, 299–307 (1970).
6. O. H. Ibarra and J. Su, "Counter machines: decision problems and applications," in: Jewels are Forever, Springer, Berlin (1999), pp. 84–96.
7. L. P. Lisovik, "On solvable problems for transducers with finite-rotary counters," Kibernetika, No. 3, 1–8 (1985).
8. Monien, "Deterministic two-way one-head pushdown automata are very powerful," Inform. Proc. Letters, **18**, 239–242 (1984).