Kybernetika

Jaroslav Král
Set-theoretical operations on k-multiple languages

Kybernetika, Vol. 3 (1967), No. 4, (315)--320

Persistent URL: http://dml.cz/dmlcz/124621

KYBERNETIKA ČÍSLO 4, ROČNÍK 3/1967

Set-Theoretical Operations on k-multiple Languages

JAROSLAV KRÁL

It is shown that the class of k-multiple languages (see [1]) is closed under formation of finite unions and intersections. The two types of complements are k-multiple modulo e. The class of k-multiple modulo e languages is closed under the formation of finite unions, not, however, under formation of intersections and complements.

The k-multiple automaton was introduced in [1] as a generalization of the concept of finite automaton and as a device for the recognition of the so called k-multiple languages. For our purposes we reformulate here some definitions from [1].

Definition 1 (Čulík). The k-multiple automaton A is defined by the $(k + 4)$-tuple $\langle V^{(1)}, V^{(2)}, \ldots, V^{(k)}, I, \Phi, i_0, F \rangle$ where

$V^{(i)}$, $i = 1, 2, \ldots, k$, are finite nonvoid sets called alphabets, elements of $V^{(i)}$ are called symbols;

I is a finite nonvoid set called the set of internal states of A;

Φ, the transition function, is a transformation from $I \otimes V^{(1)} \otimes \ldots \otimes V^{(k)}$ into I, $\otimes$ denotes the cartesian product;

i_0, the initial state, is an element of I;

F, the set of final states, is a subset of I.

A is a device which can be in some internal state $i \in I$. This device has k inputs. After reading $v_1, \ldots, v_k$ by inputs of A, the internal state i of A is changed to i_1, $i_1 = \Phi(i, v_1, v_2, \ldots, v_k)$. A can be therefore interpreted as a finite automaton with k inputs instead of one.

Definition 2. We say that a string

$$x = x_1 x_2 \ldots x_s x_{s+1} \ldots x_{2s} \ldots x_{ks}$$

is acceptable by a k-multiple automaton A if the expression

$$\Phi(\Phi(\ldots \Phi(\Phi(i_0, w_1), w_2) \ldots), w_s),$$

where $w_i = (x_i, x_{s+i}, x_{2s+i}, \ldots, x_{(k-1)s+i})$, has a meaning and defines some state from F. The string, the length of which is not the multiple of k, is not acceptable by the definition.

For a k-multiple automaton A and for an k-tuple x of symbols we shall use the terms such as "x is read by A", "x puts A into state i" and so on in the similar sense as for a finite automaton.

Definition 3. k-multiple language L_k is a set of all strings which are acceptable by some k-multiple automaton A. The automaton A will be called the automaton of L_k.

Theorem 1. *Intersection or union of two k-multiple languages is a k-multiple language.*

This is proved by a slight modification of the proof that the union or intersection of two regular events is a regular event again; see [2] or [6].

Definition 4. Complement $\tilde{L}_k$ of the k-multiple language L_k is the set

$$\tilde{L}_k = \bar{V}^* - L_k\,,$$

where $\bar{V}^*$ is the set of all strings over $\bar{V} = V^{(1)} \cup V^{(2)} \cup \ldots \cup V^{(k)}$.

Example 1. Set $L_2 = \{a^n b^n;\ n \geqq 0\}$ is the two-multiple language (see [1]). But

$$\tilde{L}_2 = \{a, b\}^* - L_2$$

and L_2 contains the set $\{a^n;\ n > 0\}$, i.e. the strings the lenghts of which are not even and we have at once:

Corollary 1. Complement of the k-multiple language L_k is not necessarily a k-multiple language.

Definition 4a. The component complement $\hat{L}_k$ of the k-multiple language L_k is the set of all strings $x \notin L_k$ of the form $d_1 d_2 \ldots d_k$, $d_i \in V^{(i)*}$ for $i = 1, 2, 3, \ldots, k$.

Henceforward in this paper by $A = \langle V^{(1)}, \ldots, V^{(k)}, I, i_0, F \rangle$ an automaton of L_k will be denoted.

Example 2. $\hat{L}_2 = \{a^n b^m;\ m \neq n;\ m, n \geqq 0\}$ is component complement of $L_2 = \{a^n b^n;\ n > 0\}$ and it follows.

Corollary 2. Component complement $\hat{L}_k$ of k-multiple language L_k is not necessarily a k-multiple language.

Definition 5. Let $V^{(1)}, \ldots, V^{(k)}$ be alphabets not containing e. A set L_k of the strings of the form $d_1 d_2 \ldots d_k$, $d_i \in V^{(i)*}$, $i = 1, 2, \ldots, k$, is a k-multiple modulo e language if and only if there exists a k-multiple language L'_k with alphabets $V^{(i)} \cup \{e\}$ so that for every $x \in L_k$ there is a $y \in L'_k$ for which $x = y \pmod{e}$ (i.e. x is equal to the y in the sense of a free semigroup with the identity symbol e generating y) and vice versa for every $y \in L_k'$ there exists $x \in L_k$ so that $y = x(\bmod e)$. In other words L_k is k-multiple modulo e if every string of L_k belongs to a k-multiple language L_k if a suitable insertion of e's is done and vice versa by erasing e's in arbitrary $y \in L_k'$ a string $x \in L_k$ is obtained.

Theorem 2. *$\tilde{L}_k$ is a k-multiple modulo e language.*

Proof. We shall construct a k-multiple automaton

$$A^0 = \langle V^0, V^0, \ldots, V^0, I^0, \Phi^0, i_0^0, F^0 \rangle, \quad V^0 = \bar{V} \cup \{e\}$$

which accepts $\tilde{L}_k$. Each string $x \in L_k$ is expressible in the form

$$x = d_1 d_2 d_3 \ldots d_k \tag{2.1}$$

where d_i are strings over $\bar{V} = V^{(1)} \cup V^{(2)} \cup \ldots \cup V^{(k)}$ and if x has the length $sk + j$, $j < k$ then $d_1, d_2, \ldots, d_j$ have the length $s + 1$ and $d_{j+1}, \ldots, d_k$ have the length s. We shall construct A^0 so that A^0 accepts only the strings x of the form $(i = 0, 1, 2, 3, \ldots)$:

$$x^0 = d_1 e^i d_2 e^i \ldots d_j e^i d_{j+1} e^{i+1} \ldots d_k e^{i+1}, \tag{2.2}$$

where $e^{i+1} = e^i e$, $i > 0$, e^0 is an empty string and d_i has the same meaning as in (2.1). It follows that the alphabets $V^{(i)}$ of A^0 are for all $i = 1, 2, \ldots, k$ equall to $V^0 = \bar{V} \cup \{e\}$. The construction of Φ^0, I^0 and F^0 is now straightforward although rather cumbersome.

If an automaton A of L_k is given by $\langle V^{(1)}, \ldots, V^{(k)}, I, \Phi, i_0 F \rangle$ we put $i_0^0 = i_0$,

$$I^0 = I \cup \{i_w^*; w = 2, 3, \ldots, k-1\} \cup \{i_D\} \cup \{i_l\}$$

where all i_w^*, $w = 2, 3, \ldots, k-1$, i_l do not belong to I. Φ^0 coincides with Φ on $I \otimes V^{(1)} \otimes \ldots \otimes V^{(k)}$. $\Phi^0(i, v_1, v_2, \ldots, v_k) = i_l$ for $v_1, v_2, \ldots, v_k \neq e$ and either $i = i_l$ or $i \in I$ and $\Phi^0(i, v_1, \ldots, v_k)$ is undefined, i.e. A^0 is in the state i_l if a symbol not belonging to $V^{(i)}$ has already been read by i-th input and the symbol e has not been read yet.

$\Phi^0(i, v_1, \ldots, v_w, e, e, \ldots, e) = i_w^*$ for $w = 2, 3, \ldots, k-1$ and $i \in I$ or $i = i_l$ (i.e. the reading of the last but one k-tuple of symbols is realized);

$\Phi^0(i, e, e, \ldots, e) = i$ for all $i \in I^0$ (i.e. reading of $(e, e, \ldots, e)$ causes no change of the internal state of A^0).

In all other cases $\Phi^0(i, v_1, v_2, \ldots, v_k) = i_D$.

Putting $i_0^0 = i_0$ and

$$F^0 = (I - F) \cup \{i_l\} \cup \{i_w^*; w = 2, 3, \ldots, k-1\}$$

we see that A^0 has all desired properties.

Theorem 3. *$\hat{L}_k$ is a k-multiple modulo e language.*

Proof. We shall construct a k-multiple automaton

$$A^c = \langle V^0, V^0, \ldots, V^0, I^c, \Phi^c, i_0^c, F^c \rangle$$

which accepts $\hat{L}_k$. (For the meaning of V^0 see the proof of the previous theorem.)

First we shall construct a k-multiple automaton $\bar{A}$ which accepts the set L_k^{ord} of strings being expressible in the form

$$x = d_1 d_2 \ldots d_k\,,$$

d_i is a string over $V^{(i)}$ for $i = 1, 2, \ldots, k$. Let

$$\bar{A} = \langle V^0, V^0, \ldots, V^0, \bar{I}, \bar{\Phi}, \bar{i}_0, \bar{F} \rangle \tag{3.1}$$

$\bar{A}$ is constructed in order to accept only the strings of the form (2.2). The construction of $\bar{A}$ is a simple matter if alphabets $V^{(i)}$ are mutually disjoint or if all $V^{(i)}$ coincide. In the general case the construction is more difficult. As the construction of $\bar{A}$ is rather cumbersome its main ideas will only be indicated. All alphabets of $\bar{A}$ are identical and equal to V^0. If $x \in L_k^{\text{ord}}$ is expressed in the form $x = d_1 d_2 \ldots d_k$ where the lenghts of d_i are s or $s + 1$, then $x^j = d_1^j d_2^j \ldots d_k^j \in L_k^{\text{ord}}$ for $j = 1, 2, \ldots, s + 1$, where (as well as below) d_i^j denotes the string formed by the first j symbols of d_i. If follows that after reading x^j there exists a finite set B_j of vectors $b = (b_i, \ldots, b_k)$ where $b_i = q_i$ if symbols from $V^{(q_i)}$ can be read by the i-th input, $i = 1, 2, \ldots, k$, so that

$$x^{j+1} = d_1^{j+1} d_2^{j+1} \ldots d_k^{j+1}$$

remains a member of L_k^{ord}.

Obviously B_{j+1} having the same meaning for x^{j+1} as B_j for x^j is a subset of B_j. Now let I contain the states of the form i_B where B is one of the above mentioned sets. Let $\Phi(i_{B_j}, V_1, \ldots, V_k) = i_{B_{j+i}}$ where $x^{j+1} = d_1^j v_1 d_2^j v_2 \ldots d_k^j v_k$, B_j containing a vector $t = (t_1, \ldots, t_n)$ so that $v_i \in V^{(t_i)}$ for $i = 1, 2, \ldots, k$. We note that these relations have a meaning as B_{j+1} is uniquely determined by B_j and $v_1, v_2, \ldots, v_k$. If B_j does not contain any vector of such a property some "absorbent" state i_D is reached i.e. for i_D it is true that $\Phi(i_D, v_1, v_2, \ldots, v_k) = i_D$ for all $(v_1, v_2, \ldots, v_k)$. The set of all i_B is finite and it can be shown that adding some auxiliary states and putting $\bar{i}_0 = i_{B_0}$, $B_0 = \{(t_1, t_2, \ldots, t_k);\ 1 \leqq t_1 \leqq t_2 \leqq \ldots \leqq t_k \leqq k\}$ it is possible to construct $\bar{A}$ of all desired properties.

Let us now construct the automaton A^c. The set of its states is formed by the set of pairs of the form $\langle i_1, i_2 \rangle$ where $i_1 \in \bar{I}$ and $i_2 \in I$ and by some additional states (i.e. the states of A^c are „pairs of states" of $\bar{A}$ and an automaton A of L_k and some aditional states).

Let $\mathbf{v} = (v_1, v_2, \ldots, v_k)$ and

$$\Phi^c(\langle i_1, i_2 \rangle, \mathbf{v}) = \langle \bar{\Phi}(i_1, \mathbf{v}), \Phi(i_2, \mathbf{v}) \rangle \tag{3.2}$$

if both Φ and $\bar{\Phi}$ are defined;

$$\Phi^c(i, e, e, \ldots, e) = i \tag{3.3}$$

for all $i \in I^c$

$$\Phi^c(\langle i_1, i_2 \rangle, \mathbf{v}) = \langle \overline{\Phi}(i_1, \mathbf{v}), i_F \rangle \tag{3.4}$$

if $\Phi(i_2, \mathbf{v})$ is not defined;

$$F^c = \{\langle i_1, i_2 \rangle; i_1 \in \overline{F}, i_2 \notin F\} \cup \{\langle i_1, i_F \rangle; i_1 \in \overline{F}\} . \tag{3.5}$$

It is easily seen that A^c has the desired properties as a state from F^c cannot be reached if $x \in L_k$ (see (3.3) and (3.4)) or if x is not expressible in the form $d_1 d_2 \ldots d_k$ where d_i is a string over $V^{(i)}$ for $i = 1, 2, \ldots, k$ (see properties of $\overline{A}$).

Theorem 5. *The union of two k-multiple modulo e languages is a k-multiple modulo e language.*

The proof is similar to the proof of the theorem 1. The only difference is that instead of considering strings x we consider the strings x' obtaining from x by convenient insertion of e's.

Example 3. Let us have two-multiple modulo e languages:

$$L_1 = \{a^n b^n c^m; m, n > 0\}$$

which is accepted by the two multiple automaton $\langle \{a\} \cup \{e\}, \{b, c, e\}, \{S_1, S_2, S_3\}, \Phi, S_1, \{S_1, S_2\} \rangle$ where $\Phi(S_1, a, b) = S_1$, $\Phi(S_1, e, c) = \Phi(S_2, e, c) = S_2$, $\Phi(S, e, c) = S$ for all S, $\Phi(\ ,\ ,\ ) = S_3$ in all other cases and

$$L_2 = \{a^m b^n c^n; m, n > 0\}$$

which is accepted by the similar automaton. But then

$$L_1 \cap L_2 = \{a^n b^n c^n; n > 0\}$$

is a three-multiple modulo e language, not a two-multiple modulo e language. It follows

Corollary 3. The intersection of two languages which are k-multiple modulo e is not necessarily a k-multiple modulo e language.

Corollary 4. The complement of k-multiple modulo e language is not necessarily a k-multiple modulo e. By the complement of L_k we mean the set

$$\tilde{L}_k = C - L_k$$

where C is the set of all strings over $\overline{V}$.

Proof. We note that for every two sets A, B

$$A \cap B = (A^c \cup B^c)^c ,$$

where ($)^c$ denotes the complement and that the assertion of the theorem follows from corollary 3 and theorem 5.

Corollary 5. The component complement $\hat{L}_k$ of k-multiple modulo e language, i.e. the set

$$\hat{L}_k = C - L_k,$$

where $C = \{d; d = d_1 d_2 \dots d_k, d_i$ is for $i = 1, 2, \dots, k$ a string over $V^{(i)}\}$ is not necessarily a k-multiple modulo e language.

The proof is the same as the proof of the previous corollary.

(Received June 1st, 1966.)

REFERENCES

[1] K. Čulík, I. Havel: On multiple finite automata. (In print.)
[2] N. Chomsky: Chapters 11—13 in Handbook of Math. Psychology. John Wiley 1963.
[3] N. Chomsky, M. P. Schützenberger: The algebraic theory of context-free languages. In Computer programming and Formal systems. North—Holland 1963.
[4] K. Čulík: Some notes on finite state languages and events represented by finite automata using labelled graphs. Časopis pro pěstování matematiky *86* (1961), 1, 43—55.
[5] S. Ginsburg, J. S. Ullian: Ambiguity in contex free languages. J. of ACM *13* (1966), 1, 62—89.
[6] В. Н. Глушков: Синтез цифровых автоматов. Физматгиз, Москва 1962.
[7] E. F. Moore: Gedankenexperimente on sequential machines. In Automata studies, Princeton 1956.
[8] C. C. Elgot, J. E. Mezei: On relations defined by generalized automata. IBM J. of Res. and Develop *9* (1965), 1, 47—68.

VÝTAH

Množinové operace nad k-násobnými jazyky

JAROSLAV KRÁL

V článku jsou zkoumány tak zvané násobné jazyky tj. jazyky akceptovatelné tzv. násobnými automaty (viz [1]), jež jsou zobecněním tzv. regulárních výrazů. Je dokázáno, že třída násobných jazyků je uzavřena vůči průniku a sjednocení, ale nikoliv vůči doplňku. Třída k-násobných modulo e jazyků je uzavřena vůči sjednocení, ale nikoliv vůči průniku a tedy ani doplňku.

Jaroslav Král, prom. matematik, Ústav výpočtové techniky ČSAV-ČVUT, Horská 3, Praha 2.