Dokl. Akad. Nauk SSSR
Tom 194 (1970), No. 6

Soviet Math. Dokl.
Vol. 11 (1970), No. 5

ESTIMATES OF THE NUMBER OF STATES OF FINITE AUTOMATA

UDC 519.95

A. N. MASLOV

In the present note we consider operations on sets of words representable in finite automata. An important measure of the complexity of these sets is the number of states of a minimal representing automaton.

Kleene [1,2] proved that a set of words is representable in a finite automaton if and only if it is obtainable from $\{\Lambda\}$ and $\{\sigma_i\}$ (where Λ is the empty word and the σ_i are the letters of the input alphabet Σ) by applying the operations of union ($\bigcup$), product $(\cdot)$, and iteration $(*)$.

In addition, $S_i x = S_k$ means that, upon receiving the word x as an input, the automaton, finding itself in state S_i, goes over into state S_k. The index 0 corresponds to the initial state if nothing is said to the contrary. $[x^\partial$ denotes the length of the word x. $T(A)$ denotes the set of words representable in the automaton A. In what follows, by representability of an event we shall always understand representability in a finite automaton.

It is well known that, if $T(A)$ and $T(B)$ are representable in automata A and B with m and n states, respectively $(m \geq 1,\ n \geq 1)$, then:

1) $T(A) \bigcup T(B)$ is representable in an automaton with $m \cdot n$ states;

2) $T(A) \cdot T(B)$ is representable in an automaton with $(m-1)\cdot 2^n + 2^{n-1}$ states $(n \geq 3)$;

3) $T(A)^*$ is representable in an automaton with $(3/4)\cdot 2^m - 1$ states $(m \geq 2)$.

Let us construct examples of automata over the alphabet $\Sigma = \{0, 1\}$ for which these estimates are attained.

1. Union. A has states $\{S_0, \cdots, S_{m-1}\}$ and transitions $S_{m-1}1 = S_0$, $S_i 1 = S_{i+1}$ for $i \neq m-1$, $S_i 0 = S_i$, and S_{m-1} is the terminal state. B has states $\{P_0, \cdots, P_{n-1}\}$ and transitions $P_i 1 = P_i$, $P_{n-1}0 = P_0$; $P_i 0 = P_{i+1}$ for $i \neq n-1$, P_{n-1} is the terminal state.

2. Product. B has the states $\{P_0, \cdots, P_{n-1}\}$ and transitions $P_{n-1}1 = P_{n-2}$, $P_{n-2}1 = P_{n-1}$, $P_i 1 = P_i$ for $i < n-2$, $P_{n-1}0 = P_{n-1}$, $P_i 0 = P_{i+1}$ for $i \neq n-1$. P_{n-1} is the terminal state. The automaton A is the same as in the case of the union.

3. Iteration. A has the states $\{S_0, \cdots, S_{m-1}\}$ and transitions $S_{m-1}1 = S_0$, $S_i 1 = S_{i+1}$ for $i \neq m-1$, $S_0 0 = S_0$, $S_i 0 = S_{i-1}$ for $i > 0$. S_{m-1} is the terminal state.

Corresponding to A and B we construct automata as in [2,4] and we find the required number of attainable and distinct states, which proves the minimality [3].

A general formulation of the problem is as follows: We have events $T(A_i)$ $(1 \leq i \leq k)$ representable in automata A_i with n_i states, respectively, and a k-place operation f on events, preserving representability in finite automata. What is the maximal number of states of a minimal automaton representing $f(T(A_1), \cdots, T(A_k))$, for the given n_i?

The problems considered above belong to this class. The result has already been obtained [2,6] that the inversion of the words of a set representable in an automaton with m states is representable in an automaton with 2^m states, and this estimate is attained over the alphabet $\Sigma = \{a, b, c\}$.

From the results of [8] it follows that, for a set T (representable in an automaton A with n states), $\{xz \mid \exists y\,(xy \cdot z \in T \ \&\ [x^{\partial} = [y^{\partial} = [z^{\partial})\}$ can be nonrepresentable. We shall prove that the sets

$$\frac{p}{q}T = \left\{x \mid \exists y \left(x \cdot y \in T \,\&\, \frac{[x^{\partial}}{[y^{\partial}} = \frac{p}{q}\right)\right\} \text{ and } \sqrt{T} = \{x \mid xx \in T\}$$

are always representable.

4. For the representation of $(p/q)\,T$, we first construct a nondeterministic automaton A_1, the states of which are the states of A, and whose transition matrix for all input letters is constructed as follows: $a_{ij} = 1 \leftrightarrow \exists x\,([x^{\partial} = p \ \&\ S_j x = S_i)$. The set of initial states P_0 consists of the terminal states of A. From [5] it is known that one can obtain a deterministic automaton A_2 having not more than $2^{c\sqrt{n \ln n}}$ attainable states. The desired automaton has states $(S_i|l|P_j)$, where S_i is a state of A, P_j is a state of A_2, $0 \le l < q$ is an integer. The transitions are defined by the formula: $(S_i|l|P_j)\,\sigma = (S_i\sigma|l + 1 \pmod q|P_j\sigma^{1-\operatorname{sign} l})$. $(S_0|0|P_0)$ is the initial state. A state is terminal if $l = 0$ and $S_i \in P_j$. The number $N(p, q, n)$ of states of this automaton is $\le qn2^{c\sqrt{n \ln n}}$. A lower bound for $\log_2 N(p, q, n)$ can be obtained using examples in the paper [5], but this will differ by a multiplicative constant.

In addition, we shall need an estimate of the number of congruence classes of a representable event [2].

5. To each word Z there corresponds a mapping of states Z: $S_i \to S_i Z$. It is obvious that, if the words x and y have the same corresponding mappings, then they are congruent, i.e. there are not more than n^n congruence classes. We consider an automaton from the paper [7]. This is an automaton B^n over the alphabet $\{a, b, c\}$ with states $\{Q_0, \cdots, Q_{n-1}\}$ and transitions $Q_{n-1}a = Q_0$, $Q_i a = Q_{i+1}$ for $i \ne n-1$, $Q_0 b = Q_1$, $Q_1 b = Q_0$, $Q_i b = Q_i$ for $i > 1$, $Q_0 c = Q_0$, $Q_1 c = Q_0$, $Q_i c = Q_i$ for $i > 1$. Q_0 is terminal. B^n has exactly n^n congruence classes.

6. Over the alphabet $\Sigma = \{0, 1\}$ there cannot exist automata with n states and n^n congruence classes. The precise value of the maximum number of congruence classes in this case is unknown. However, one can construct an automaton B_1^n having not less than $(n-1)^{n-1}$. Its states are $\{P_0, \cdots, P_{n-1}\}$ and its transitions are: $P_{n-1}1 = P_1$, $P_0 1 = P_2$, $P_i 1 = P_{i+1}$ for $i \ne 0, n-1$, $P_i 0 = P_0$, $P_0 0 = P_2$, $P_2 0 = P_1$, $P_i 0 = P_i$ for $i > 2$. P_1 is simultaneously an initial and terminal state.

We remark that the states $P_1, \cdots, P_{n-1}$ are mapped into each other by the words $a' = 1$, $b' = 001^{n-1}$ and $c' = 01^{n-1}$, in the same way as the states of the automaton B^{n-1} by the words a, b, c. Hence, in the set $(a' \cup b' \cup c')^*$ there are $(n-1)^{n-1}$ noncongruent words.

7. The set $\sqrt{T}$ is representable in an automaton the states of which are ordered sequences of length n of states of A; the transitions are given by the formula

$$(S_{i_1}, \dots, S_{i_n})\,\sigma = (S_{i_1}\sigma, \dots, S_{i_n}\sigma),$$

$(S_0, S_1, \cdots, S_{n-1})$ is the initial state. A state is terminal if the sequence has a coordinate $S_{i_{l_1}}$ which is a terminal state of A. The automata B^n and B_1^n give the lower bounds n^n and $(n-1)^{n-1}$.

8. In [4] it was shown that permutations of letters in words can lead from representable sets to nonrepresentable sets. If one permits only cyclical premutations, i.e.

$$T' = \{x \mid x = \sigma_{i_1} \ldots \sigma_{i_k} \& \exists l \, (\sigma_{i_l} \ldots \sigma_{i_k} \sigma_{i_1} \ldots \sigma_{i_{l-1}} \in T)\}$$

(where T is representable in an automaton A with n states), then one can prove that T' is representable in an automaton B with $(n2^n - 2^{n-1})^n$ states. In fact, let B_i have the same states and transitions as A, let S_i be its initial state, and let its terminal states be the same as in A. Let C_i have the same states and transitions as A, let S_0 be its initial state and S_i its only terminal state. $T' = \bigcup_{i=0}^{n-1}(T(B_i) \cdot T(C_i))$ and, therefore, (cf. 1) and 2)) it is representable in an automaton with $(n2^n - 2^{n-1})^n$ states. Below we indicate an example giving a lower bound $((n-2)2^{n-2})^{n-2}$ for $n > 3$. Of course, the input alphabet increases as n increases. $\Sigma = \{a_0, \cdots, a_{n-2}, b_0, \cdots, b_{n-2}\}$. The states of the automaton are $\{S_0, \cdots, S_{n-1}\}$, and its transitions are: $S_i a_i = S_{n-1}$, $S_{n-1} a_i = S_i$, $S_j a_i = S_j$, for $i \neq j$, $S_0 b_i = S_i$, $S_j b_i = S_j$ for $i \neq j$. The terminal state is S_{n-1}.

The author expresses gratitude to A. A. Mučnik for his guidance.

Moscow State University

Received 18/MAR/70

BIBLIOGRAPHY

[1] S. C. Kleene, *Representation of events in nerve nets and finite automata*, Automata Studies, Ann. of Math. Studies, no. 34, Princeton Univ. Press, Princeton, N. J., 1956, pp. 3–41; Russian transl., Automata, Moscow, 1956, pp. 15–67. MR 17, 1040.

[2] M. O. Rabin and D. Scott, *Finite automata and their decision problem*, IBM J. Res. Develop. 3 (1959), 114–125; Russian transl., Kibernet. Sb. 4 (1962), 58. MR 21 #2559.

[3] E. F. Moore, *Gedanken-experiments on sequential machines*, Automata Studies, Ann. of Math. Studies, no. 34, Princeton Univ. Press, Princeton, N. J., 1956, pp. 129–153; Russian transl., Automata, Moscow, 1956. MR 17, 1140.

[4] V. M. Gluškov, *Abstract theory of automata*, Uspehi Mat. Nauk 16 (1961), no. 5 (101), 3–62 = Russian Math. Surveys 16 (1961), no. 5, 1–53. MR 25 #1976.

[5] Ju. I. Ljubič, *Estimates for the optimal determination of indeterminate autonomous automata*, Sibirsk. Mat. Ž. 5 (1964), 337–355. (Russian) MR 29 #29.

[6] B. G. Mirkin, *On dual automata*, Kibernetika (Kiev) 1966, no. 1, 7–10. (Russian) MR 33 #7209.

[7] A. Paz, *Homomorphisms between finite automata*, Bull. Res. Council Israel Sect. F 10F (1962), 93–100. MR 26 #4865.

[8] R. E. Stearns and J. Hartmanis, *Regularity preserving modifications on regular expressions*, Information and Control 6 (1963), 55–69. MR 28 #3905.

Translated by:
E. Mendelson