An approach to the classification of the loops of finite automata. 
Part I: Long corresponding loops

B. F. Melnikov, A. A. Melnikova

Abstract—In this paper, we considered questions of the possible classification of the states and loops of a nondeterministic finite automaton. For the development of algorithms for equivalent transformation of nondeterministic finite automata, we consider the basis finite automaton for the given regular language and the paths and loops of its transition graph. We also consider the paths and loops of the transition graph of another nondeterministic automaton that defines the same language. On the basis of this, we define corresponding paths and loops of two mentioned automata and the questions of their classification. This classification gives, for example, the possibilities for describing some heuristic algorithms for minimization of nondeterministic automata.

For the last thing, we describe the following objects. For each state of the basis automaton, we consider the states of the given automaton corresponding to this state of the basis automaton, and give their classification as a function of the loops passing through the same state of the basis automaton. Their subset is the set of so-called including loops, on the basis of which we determine so-called partially complete loops. For any chosen vertex of the basic automaton, we call the vertices of the considered nondeterministic automaton, through which all possible partially complete loops pass, by complete cyclic states.

At the end of the paper, we formulate the hypothesis that if for any state of the considered nondeterministic automaton, there exists at least one corresponding state of the basis automaton, such that the first one is a complete cyclic state for the second one, then all the corresponding states of the basis automaton are such ones.

In the presented Part I of the paper, we consider the issues related to corresponding loops of the given nondeterministic automaton and equivalent basis automaton.

Keywords—nondeterministic finite automata, basis automaton, transition graph, path, loop, algorithms for equivalent transformation, universal automaton.

I. INTRODUCTION AND MOTIVATION

We continue the series of papers of equivalent transformation of nondeterministic finite automata (NFAs); let us mention some of them: [1]–[7]. One (but not the only) goal of such publications is the description of algorithms for NFA-minimization. Speaking of this, we mean simultaneously some following dichotomies (trichotomies, etc.).

- The search for an equivalent automaton with a minimum number:
  - of edges, see [2], [4] etc.;
  - or of the star-height value, see [10], [11] etc.

More than that, it is possible to minimize by some mixed criteria; we are going to consider such options for minimization in one of the following publications.

- We consider both exact and heuristic algorithms; and among the last ones, we consider:
  - “usual” heuristic algorithms;
  - and so-called anytime algorithms, [12] etc.

It is important to note, that for all resulting classes of algorithms, we use the same auxiliary algorithms of equivalent transformation of NFA.

In this paper, we continue to consider the basis finite automaton for the given regular language, and also the paths and loops of its transition graph. We also consider the corresponding paths of the transition graph of some NFA that defines the given language.

Considered in this paper questions of the possible classification of the states and loops of a NFA give additional possibilities for describing algorithms mentioned before. For this thing, we consistently describe the following objects.

- For each state of the basis automaton, we consider the states of the given automaton corresponding to this state of the basis automaton. We give their classification as a function of the loops passing through them corresponding to the same state of the basis automaton.
- Then we formulate two special properties of the automaton $K$ related to all their states; for different NFAs defining the same regular language, these properties can either be fulfilled or not fulfilled.
- The special subset of loops of the given automaton corresponding to the same state of the basis automaton is the set of so-called including loops.
- On the basis of the last definition, we determine so-called partially complete loops.
- For any chosen vertex of the basic automaton, we call the vertices of the considered nondeterministic automaton through which all possible partially complete loops pass, by complete cyclic states.
- At the end of the paper, we formulate the following hypothesis. If for any state of the considered NFA, there exists at least one corresponding state of the basic automaton, such that the first one is a complete cyclic state for the second one, then all the corresponding states of the basic automaton are such ones.

These objects and hypothesis are the main subject of this paper.

Received August 14, 2018.
Boris F. Melnikov, Russian State Social University (email: bf-melnikov@yandex.ru).
Aleksandra A. Melnikova, National Research Nuclear University “MEPhI” (email: super-avahi@yandex.ru).
Note that in the works cited above, we also give the algorithms for constructing all these objects; certainly, the algorithms are based on the regular language $L$ only. Some examples of the defined objects were given in \cite{3, 7}. Let us consider one of examples once again. Let us consider the language accepted either by automaton given on Fig. 1 ...

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig1.png}
\caption{Fig. 1}
\end{figure}

... or by equivalent automaton in canonical form given on Fig. 2.

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig2.png}
\caption{Fig. 2}
\end{figure}

And the following Fig. 3 and Table 1 (see also \cite{4, Tab.12}) show corresponding automaton $L^R$ ...

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{fig3.png}
\caption{Fig. 3}
\end{figure}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\# & X & Y & Z & U \\
\hline
A & – & # & – & – \\
B & # & – & # & – \\
C & # & # & # & – \\
D & # & # & # & – \\
\hline
\end{tabular}
\caption{Tab. 1}
\end{table}

We also obtain the following automaton $\mathcal{B}_4(L)$, see Tab. 2 below. Remark that it was not considered in \cite{4}, and was considered in \cite{7}; we repeat the table of the last paper. Also, for the usability, we write $A \# X$, $A \# Y$ instead of $\{A, X, Y\}$, etc. Also we do not write symbols of sets (braces), i.e., we write $A \# X, A \# Y$ instead of $\{A, X, Y\}$, etc.
III. PATHS OF THE BASIS AUTOMATON AND CORRESPONDING PATHS OF AN EQUIVALENT AUTOMATON

In this section and following ones, we consider a “natural” correspondence between loops of some automaton \( \hat{L} \) defining the given regular language \( L \) and equivalent basis automaton \( \hat{B}(L) \). As we said before, a part of the material in this section was already published in Russian in [6].

For the considered language \( L \) and its basis automaton \( \hat{B}(L) \) \( \hat{2} \), let us consider the transition function \( \hat{\delta} \) of \( \hat{B}(L) \) as the set of its edges

\[
\hat{\Delta} = \{ \hat{e}_i = (\hat{p}_i, a_i, \hat{r}_i) \mid \\
\hat{p}_i, \hat{r}_i \in \hat{Q}, a_i \in \Sigma, \delta(\hat{p}_i, a_i) \ni \hat{r}_i, i \in \{1, \ldots, \hat{m}\}\}
\]

(where \( \hat{m} \) is the number of edges of automaton \( \hat{2} \)). Thus, we have numbered (“colored”) the edges of automaton \( \hat{B}(L) \).

And the transition function of automaton \( \hat{1} \) can also be considered as the set of edges

\[
\Delta = \{ e_j=(p_j, a_j, r_j) \mid \\
p_j, r_j \in Q, a_j \in \Sigma, \delta(p_j, a_j) \ni r_j, j \in \{1, \ldots, m\}\}
\]

(where \( m \) is the number of edges of automaton \( \hat{1} \)).

Let us mark each edge of automaton \( \hat{1} \) by a set of possible (already used) “colors”, i.e., let us consider function

\[
\Omega: \Delta \to \mathcal{P}(\hat{\Delta})
\]

crafted in the following way. For the edge

\[
e_j = (p_j, a_j, r_j),
\]

we set \( \Omega(e_j) \ni \hat{e}_i \) (where \( i \in \{1, \ldots, \hat{m}\} \)) if and only if the following conditions hold:

\[
[p_j \ni \hat{p}_i]; \quad a_j = a_i; \quad [r_j \ni \hat{r}_i].
\]

Let us consider the set of paths of automaton \( \hat{B}(L) \); it is important to remark, that we consider not only simple paths. We can denote a path of automaton \( \hat{2} \) by \( \hat{\nu} \) and consider it as the sequence

\[
\hat{\nu} = (\hat{e}_{\hat{n}}, \ldots, \hat{e}_1),
\]

where \( \hat{n} \) is the number of its edges. Its \( k \)-th edge \( \hat{e}_k \) will be denoted by

\[
\hat{e}_k = (\hat{p}_k, a_k, \hat{r}_k)\]

\((k \in \{1, \ldots, \hat{n}\})\). Certainly, \( \hat{p}_{k+1} \ni \hat{r}_k \) for each possible \( k \).

For the path \( \hat{3} \), let us define the set of \( \hat{\nu} \)-corresponding paths of automaton \( \hat{K} \). Thus, consider a path

\[
\nu = (e_{\nu}, \ldots, e_0)
\]

of automaton \( \hat{K} \). The value \( \hat{n} \) is the same as in \( \hat{3} \) (remember that both \( \hat{K} \) and \( \hat{B}(L) \) do not contain \( \varepsilon \)-edges), and its \( k \)-th edge is

\[
e'_k = (p'_k, a'_k, r'_k)
\]

for each possible \( k \). Let us say, that this path corresponds to \( \hat{\nu} \), if for each \( k \in \{1, \ldots, \hat{n}\} \),

\[
\Omega(e'_k) \ni \hat{e}_k.
\]

(I.e., if each edge can be “colored by the color” of the edge of path \( \hat{\nu} \) with the same number.) We shall also say, that for such paths, the states \( \hat{p}_k \) and \( p'_k \) are corresponding for each \( k \in \{1, \ldots, \hat{n}\} \).

Let us consider some examples. Consider the language which can be defined by regular expression

\[
(a + ab + ba)^*.
\]

Let us now consider the graphical description of the following automata.

- \( \hat{L} \), see Fig. 4 below.

Remark that in this example, we have \( \hat{L} = \hat{L}^{\hat{R}} \).

Therefore considering Fig. 4 and changing marks \( A, B \) and \( C \) for \( X, Y \) and \( Z \) respectively, we obtain automaton \( \hat{L}^{\hat{R}} \).

- \( \hat{B}(L) \), see Fig. 5:
• The equivalent automaton $K$, see Fig. 6 below. Like the agreements of [7], we write here also the values of state-marking functions.

For example, let us consider only two the following edges of automaton $BA(L)$: the “green” edge

$$A \xrightarrow{a} B,$$

and the “blue” edge

$$B \xrightarrow{a} B.$$

Therefore (Fig. 6), the following edges of automaton $K$ (and only they) can be “green”:

$$q_1 \xrightarrow{a} q_1, \quad q_1 \xrightarrow{a} q_2 \quad \text{and} \quad q_3 \xrightarrow{a} q_1,$$

and the following edges (and only they) can be “blue”:

$$q_1 \xrightarrow{a} q_1 \quad \text{and} \quad q_1 \xrightarrow{a} q_2.$$

Therefore we obtain, e.g., that the edge

$$q_1 \xrightarrow{a} q_1$$

can be “colored” in both “green” and “blue” colors (and, maybe, in some other ones).

Let us also remark, that the set of simple loops of automaton $BA(L)$ includes the loop

$$A \xrightarrow{Y} B \xrightarrow{a} B \xrightarrow{a} B \xrightarrow{b} A \xrightarrow{Y} A.$$  \hfill (6)

We shall use this loop in some next examples.

IV. THE FIRST PROPERTY OF AN AUTOMATON

In the example considered before, for each state $Y$ of $BA(L)$ there was a corresponding state of $K$ containing all the corresponding loops. In other words, the following fact held in previous example.

**Property 1:** For each state $Y$ of $BA(L)$, there exists a state $q$ of $K$ (where $q \ni A \ni X$), such that for each loop $\nu$ of state $X$, there exists a loop of $K$ corresponding to $\nu$.

The next example shows, that there exist some automata where Property 1 does not hold. Thus, let us consider such example briefly, the considered automaton is given on Fig. 7:

![Fig. 7](image)

Remark that we can think here, that automaton on Fig. 7 is simultaneously:

• $\tilde{L}_L$,
• $(L^R)^R$ (changing mark $A$ for $X$);
• $BA(L)$ (changing mark $A$ for $\hat{A}$).

This automaton defines the whole language $\Sigma^*$ (we can think that $\Sigma = \{a, b\}$). And both the states of equivalent automaton $K$ on Fig. 8 . . .

![Fig. 8](image)

... have the same values of $\varphi_{in}^n$ and $\varphi_{out}^n$ (i.e., $\{A\}$ and $\{B\}$ respectively), but neither of states of $K$ has both loops corresponding to the loops of $BA(L)$ with labels $a$ and $b$.

V. LONG CORRESPONDING LOOPS

Let us define long loops. Firstly, let a path (3) defined in Section III be the loop. Then let us define its $n$-loop (i.e., long loop) for each $n \geq 1$ in the usual way.

**Definition 1:** For $n \geq 1$, the $n$-loop of the loop $\nu = (e_1^\nu, \ldots, e_n^\nu)$ is defined by

$$\nu^n = (e_1^{\nu^n}, \ldots, e_n^{\nu^n}),$$  \hfill (7)

where for each possible $i$ (i.e., $i \in \{1, \ldots, (n-1)\cdot \nu\}$) we think that

$$e_i^{\nu^n} = \tilde{e}_{i+n}^{\nu^n}.$$

For this definition, let us remark the following things. Firstly,

$$\tilde{e}_{n-n}^{\nu^n} = \tilde{p}_1^{\nu^n}.$$

Secondly, we used notation $\tilde{\cdot}, \nu, \cdot$, etc; but we can assume, that we defined such long loops not only for basis automaton, but also for each automaton (for example, for the given automaton (1)).

For such long loops, we consider corresponding paths of automaton (1) as we did in Section III. We shall call such path

$$\nu = (e_1^{\nu}, \ldots, e_n^{\nu}, n \cdot \nu),$$  \hfill (8)

by “long $n$-path” of automaton (1) corresponding to loop (2) of automaton (2).

By (15), path (8) is a loop if $\nu^{\nu} = p_i^{\nu}$; then we shall call it by (long) $n$-loop. (In fact, we have already considered some examples of such long loops in previous section.)

Let us consider once again regular language defined by (5) and its automaton $BA(L)$ on Fig. 5. We describe the possible method of duplicating the loop. Let us remark in advance,
that we can simply consider Fig. 11 only (i.e., consider it as the given automaton), proving the equivalence of that automaton and automaton on Fig. 9.

Thus, let us add the duplicate of the loop $\delta$; we obtain automaton on Fig. 9. Then using a simple operation (see Fig. 10, heavy and deleted lines there), we change:

- edge $(A Y)_1 \xrightarrow{a} (B Y)_1$ for edge $(A Y)_1 \xrightarrow{a} (B Y)_2$ (remark that we cannot use the notation $\delta$ for transition function here, therefore we simply omit the name of this function);
- edge $(A Y)_2 \xrightarrow{a} (B Y)_2$ for edge $(A Y)_2 \xrightarrow{a} (B Y)_1$, and obtain automaton on Fig. 11. (We think, that there is no need to describe such operation detailed.)

The last automaton has, e.g., the long loop

$$(A Y)_1 \xrightarrow{a} \hat{\delta} (B Y)_2 \xrightarrow{a} \hat{\delta} (B X)_2 \xrightarrow{b} \hat{\delta} (A Y)_2 \xrightarrow{a} \hat{\delta} (B Y)_1 \xrightarrow{a} \hat{\delta} (B X)_1 \xrightarrow{b} \hat{\delta} (A Y)_1$$

corresponding to the long loop

$$(A Y)_1 \xrightarrow{a} \hat{\delta} (B Y)_2 \xrightarrow{a} \hat{\delta} (B X)_2 \xrightarrow{b} \hat{\delta} (A Y)_2 \xrightarrow{a} \hat{\delta} (B Y)_1 \xrightarrow{a} \hat{\delta} (B X)_1 \xrightarrow{b} \hat{\delta} (A Y)_1$$

of the given basis automaton $BA(L)$.

We shall continue to consider the last automaton in Part II of this paper.

VI. THE SECOND PROPERTY OF AN AUTOMATON

Let us complicate the condition formulated in Section III. Thus, in the examples considered before (automata on all the figures considered before, excepting Fig. 8), the following condition held: for each state $A X$ of $BA(L)$, there was a corresponding state of $K$ containing all the corresponding long loops. (Recall that before the similar condition, i.e. Property 1, for corresponding loops only, not for long loops.)

In other words, the following fact held in all previous examples of Section V.

Property 2: For each state $A X$ of $BA(L)$, there exists a state $q$ of $K$ (where $[q \ni A X]$),
such that for each loop $\nu$ of state $A_i$, there exists a number $n \geq 1$, such that there exists a $q$-loop corresponding to $\nu^n$. □

However, the considered before automaton on Fig. 8 shows, that there exist some automata where this fact also does not hold. (As before, $K$ contains a path corresponding to $\nu^n$ for each $n \geq 1$.) And the following simplest Table 3 shows all the possibilities of fulfilling or not fulfilling the formulated properties for NFAs and corresponding examples of the automata considered before.

<table>
<thead>
<tr>
<th>Automaton</th>
<th>Property 1</th>
<th>Property 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 7</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Fig. 11</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>Fig. 8</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

In the following proposition, we consider the first of the properties of corresponding loops.

**Proposition 1:** For each loop of automaton $B(A(L)$ there exists $n \geq 1$, and for it there exists a corresponding $n$-loop of automaton $K$.

(Notice that we do not require that the given loop of automaton $B(A(L)$ is simple.)

**Proof.** Consider a loop $\nu = (e_1^\nu, \ldots, e_n^\nu)$ of automaton $B(A(L)$, starting and finishing in state $p_1^n = p_h^n$.

Let $v$ be its label ($v \neq \varepsilon$, because $B(A(L)$ has no $\varepsilon$-edges), let also

$$u \in L_{\text{in}}(B(A(L)) \left( p_1^n \right)$$

and

$$w \in L_{\text{out}}(B(A(L)) \left( p_h^n \right);$$

remark that $uwv w \in L$ for each $i \geq 0$.

Let $n$ be greater than the number of states of automaton $K$. Consider word

$$x = uv^n w \in L;$$

because automaton $B(A(L)$ is unambiguous, it accepts $x$ passing $(n$ times) loop $\nu$. Automaton $K$ is equivalent (i.e., it defines $L$), then it accepts $x$.

Consider one of possible paths for automaton $K$ defining $x$; by definition, some its sub-paths correspond to considered loop $\nu$ of automaton $B(A(L)$. For this path of $K$, consider the set of states passed before reading the first letter of $v$ in the word $x$; we have $n$ such cases of reading $v$. Because the number of states of automaton $K$ is less than $n$, we obtain the coincidence for at least 2 states of this set, and, therefore, the existence of corresponding long loop. □

We shall continue to consider long corresponding loops in Part II of this paper. Besides, we shall consider there some other properties of the states and loops of an arbitrary finite automaton, completing our approach to their classification. We also shall define so-called including loops, partially complete loops and complete cyclic states, consider some their properties and formulate an important hypothesis about them; in Introduction, we already gave brief information on these objects.

Remark in the conclusion of Part I, that in [6] we also considered cases, when a NFA defines not the given language (i.e., language of $B(A(L)$ according to the terminology used here), but some its own subset; in both parts of this paper, we shall not consider these issues.

REFERENCES


Boris Felikovich MELNIKOV,
Professor of Russian State Social University (http://www.rgsu.net/),
email: bf-melnikov@yandex.ru,
mathnet.ru: personid=27967,
elibrary.ru: authorid=15715,
scopus.com: authorID=55954040300.

Aleksandra Aleksandrovna MELNIKOV,
Associated Professor (Docent)
of National Research Nuclear University “MEPhI”
(http://mephi.ru/),
email: super-avahi@yandex.ru,
elibrary.ru: authorid=143351.