DOI: 10.18698/1812-3368-2016-2-3-15

УДК 519.7

К ВОПРОСУ ЧАСТИЧНОГО УГАДЫВАНИЯ ФОРМАЛЬНЫХ ЯЗЫКОВ

Р.С. Исмагилов, **А.А. Мастихина**

МГТУ им. Н.Э. Баумана, Москва, Российская Федерация
e-mail: anmast@bmstu.ru

Рассмотрены бесконечные цепочки символов некоторого алфавита, порожденные размеченным ориентированным графом. Модифицировано понятие частичного угадывания. Изложены методы частичного угадывания для класса языков, основанные на ином подходе к рассматриваемым задачам. Доказан критерий и приведен конструктивный алгоритм угадывания. Сопоставлены результаты, полученные в настоящей работе, с результатами, полученными ранее. Изложение замкнуто в себе и использует лишь элементарные понятия, связанные с графами и автоматами.

***Ключевые слова*:** частичное угадывание, граф, автомат.

TO THE PROBLEM OF PARTIAL GUESSING OF FORMAL LANGUAGES

R.S. Ismagilov, **A.A. Mastikhina**

Bauman Moscow State Technical University, Moscow, Russian Federation
e-mail: anmast@bmstu.ru

The purpose of the study is to examine the infinite chains of symbols of an alphabet, generated by a marked directional graph. The findings of the research helped us to modify the notion of partial guessing. For a class of languages, we employed methods of partial guessing based on a different approach to the problems under consideration. We found a criterion and a constructive algorithm of guessing. We compared the results obtained in this study with the results obtained previously. In this research we present only basic concepts associated with graphs and automata.

***Keywords*:** partial guessing, graph, automata.

Введение. Задача угадывания может быть описана (в вольном изложении) следующим образом. Имеется бесконечная цепочка символов некоторого алфавита (сверхслово), порожденная некоторым механизмом (в каждый момент времени поступает один символ). В качестве такого механизма берутся размеченные ориентированные графы (автоматы). Зная начало указанной цепочки, требуется предсказать (угадать) следующий ее символ; разумеется, такое угадывание также выполняется автоматом. Некоторые аспекты этой задачи рассмотрены в работе [1]. Достаточно естественным представляется несколько "смягчить" описанную "жесткую" постановку (угадать каждый символ

начиная с некоторого места), требуя угадать символы лишь для некоторого (желательно, достаточно большого) множества моментов их поступления. Частота моментов удачного угадывания — основная характеристика процесса. Задача частичной угадываемости языков изучена в работе [2]. В частности, в ней введена упомянутая характеристика, названная *степенью угадывания*, и дана ее оценка для некоторых естественных классов языков.

В настоящей работе, во-первых, модифицировано (упрощено) понятие частичного угадывания, во-вторых, изложены методы частичного угадывания для класса языков, рассмотренных в работе [3]. Эти методы основаны на ином подходе к таким задачам; как следствие, они проще тех, которые рассмотрены в работе [3] и легко реализуемы. Сопоставлены результаты, полученные в данной статье, с результатами, приведенными в работе [3]. В изложении использованы лишь элементарные понятия, связанные с графами и автоматами.

Основные понятия. Постановка задачи. Приведем необходимые сведения о графах и языках, а также основную задачу — угадывание языков и множеств путей в графах.

Орграфы. Рассмотрим конечные ориентированные графы (орграфы) вида $G = (Q, E)$, где Q — множество вершин; E — множество ориентированных ребер; ребра запишем в виде $a \to b$. Допустимы петли $a \to a, a \in Q$. Для любого множества $Q' \subset Q$ обозначим через $[Q']$ подграф (Q', E'), E' — множество всех ребер графа (Q, E), вершины которых принадлежат множеству Q'. В частности, граф $G = (Q, E)$ обозначим через $[Q]$, в том случае, когда это не будет вызывать неоднозначности. Граф, содержащий хотя бы одно ребро, назовем нетривиальным.

Граф называется *сильно связным*, если для любых двух его вершин a_1 и a_2 найдется путь из a_1 в a_2 и путь из a_2 в a_1. Если $G = (Q, E)$ — произвольный орграф, то множество $Q' \subset Q$ назовем сильно связным, если подграф $[Q']$ является сильно связным. Максимальный (по включению) сильно связный подграф графа назовем сильно связной компонентой. Пути (бесконечные вправо) будем записывать в виде последовательности вершин $a_1 \to a_2 \to \ldots$ Множество всех таких путей обозначим через $R(Q)$, а множество путей, исходящих из вершины q_0, — через $R(q_0, Q)$. Конечный путь $a_i \to \ldots \to a_j, j > i$, назовем *отрезком пути* $a_1 \to a_2 \to \ldots$

Определение 1. *Множество вершин $Q' \subseteq Q$, встречающихся на пути $r = a_1 \to a_2 \to \ldots$ бесконечно много раз, назовем предельным множеством данного пути.*

Используем обозначение $Q' = \overline{\lim}\, r$, или $Q' = \overline{\lim}\, a_i$.

Лемма 1. *Для любого пути $r = a_1 \to a_2 \to \ldots$ множество Q' сильно связно. Существует такой номер m, что все вершины $a_i, i > m$, лежат в множестве Q'.*

Доказательство леммы (весьма простое) опускаем. Итак, в любом орграфе любой (бесконечный вправо) путь после отбрасывания нескольких его начальных вершин лежит в сильно связном подграфе (предельном множестве пути). Этот простой факт окажется весьма полезным для основной цели настоящей работы (задачи угадывания).

Возьмем произвольное сильно связное множество $Q' \subset Q$ и обозначим через $R(Q, Q')$ множество всех путей с предельным множеством Q'. Положим $R(q_0, Q, Q') = R(Q, Q') \cap R(q_0, Q)$. Приведенная ниже лемма немедленно следует из леммы 1.

Лемма 2. *Справедливо равенство $R(Q) = \cup_{Q'} R(Q, Q')$ (объединение по всевозможным сильно связным подмножествам Q' орграфа).*

Языки и орграфы. Используем только языки в алфавите $\{0, 1\}$. Рассмотрим множество $\{0, 1\}^\infty$, состоящее из всех бесконечных последовательностей $\alpha = \alpha(1)\alpha(2)\ldots$, где $\alpha(i) \in \{0, 1\}, i = 1, 2, \ldots$; они называются *словами*. Удобно полагать, что в каждый момент времени n подается символ $\alpha(n)$ слова α. Любое подмножество $L \subset \{0, 1\}^\infty$ называется *языком* (в алфавите $\{0, 1\}$). Применяются также термины "сверхслово", "ω-слово", "сверхязык", "ω-язык". Здесь выбраны более короткие названия.

Для создания языков используем орграфы $G = (Q, E)$ со следующим свойством: из каждой вершины исходят либо два ребра, либо одно. Если из каждой вершины исходят два ребра, назовем граф *совершенным*, если имеются вершины, из которых выходит единственное ребро, то назовем граф *несовершенным*. Пусть дано отображение $f : E \to \{0, 1\}$. Величину $f(a, b)$ назовем *f-меткой* на ребре $a \to b$. Потребуем, чтобы $f(a, b) \neq f(a, c)$ в случае, когда из вершины исходят два ребра $a \to b$ и $a \to c$. Полученный объект (*размеченный орграф*) обозначим через $G = (Q, E, f)$. Если фиксирована вершина q_0, то запишем $G = (Q, E, f, q_0)$.

Каждому пути $a_1 \to a_2 \to \ldots$ поставим в соответствие слово $\alpha = \alpha(1)\alpha(2)\ldots$, где $\alpha(i) = f(a_i, a_{i+1})$, $i = 1, 2, \ldots$ Получим отображение $R(Q) \to \{0, 1\}^\infty$. При таком отображении образы множеств $R(Q), R(Q, Q')$ и $R(q_0, Q, Q')$ обозначим через $L(Q), L(Q, Q')$ и $L(q_0, Q, Q')$. В случае совершенного графа отображение $R(q_0, Q) \to$ $\to \{0, 1\}^\infty$ взаимно-однозначно.

Ограничимся графами, в которых полустепень исхода каждой вершины не превосходит двух. Это обусловлено тем, что с графом связываем язык с алфавитом из двух символов. Рассмотрение графов без

указанного свойства (при сохранении указанного алфавита) потребовало бы существенного изменения (и усложнения) теории.

Угадывание языков и путей. Проблема угадывания слова (в алфавите $\{0,1\}$) состоит (неформально) в следующем: зная первые n символов слова $\alpha \in L$ (слово $\alpha(1)\alpha(2)\dots\alpha(n)$), угадать следующий символ $\alpha(n+1)$. Обозначим через $\beta(n)$ "прогноз" относительно того, каким будет $(n+1)$-й символ $\alpha(n+1)$ слова α. Если $\alpha(n+1)=\beta(n)$ для некоторого n, то можно утверждать, что n-й символ слова α *правильно угадан* (или угадан), число n назовем *моментом правильного угадывания* (или моментом угадывания) этого символа. Запишем цепочку прогнозов:

$$\beta(1),\dots,\beta(n),\dots \tag{1}$$

Таким образом, процесс угадывания слова можно записать в виде

$$\alpha(1),\beta(1),\dots\alpha(n),\beta(n),\dots \tag{2}$$

Процесс угадывания языка L задается цепочками вида (1) для каждого слова $\alpha \in L$.

Задача угадывания пути в орграфе — зная пройденные вершины $a_1,\dots,a_k$ пути, предсказать следующую вершину a_{k+1}. Если зафиксировать вершину q_0 орграфа, то пути, выходящие из вершины q_0, отождествляются со словами в алфавите и тем самым задачи угадывания слов и путей оказываются эквивалентными.

Несколько модифицируем определение, полагая, что угадывание может происходить не в каждый момент, а в моменты времени $k_1,k_2,\dots$ Таким образом, в цепочке (1) будут отсутствовать некоторые прогнозы; условимся писать черточки на этих позициях (например, запись $0,1,-,0,\dots$ будет означать, что $\beta(1)=0,\beta(2)=1$, в момент 3 угадывания нет и т.д.).

В целях реализации угадывания путей на орграфе поставим в соответствие некоторым (возможно, не всем) ребрам $a \to b$ метки $g(a,b) \in \{0,1\}$; назовем их g-метками. Если, двигаясь по пути, дойти до вершины a_k данного пути и на ребре $a_{k-1} \to a_k$ считать g-метку $g(a_{k-1},a_k)$, то можно спрогнозировать, что следующей будет такая вершина a_{k+1}, что $f(a_k,a_{k+1})=g(a_{k-1},a_k)$. Разумеется, придется угадывать множества путей $R \subset R(Q)$, а не только отдельные пути.

Соответственно, если речь идет об угадывании слова, то прогнозируем, что следующей буквой слова будет $g(a_{k-1},a_k)$. Таким образом, указанная выше "угадывающая функция" F задается равенством $F(\alpha(1)\alpha(2)\dots\alpha(k))=g(a_{k-1},a_k)$.

Итак, роль меток на ребрах такая: f-метки позволяют установить биекцию между словами и путями (в этом месте рассматриваются

пути, выходящие из фиксированной вершины q_0); g-метки дают прогноз относительно очередной буквы слова (при условии, что известны предшествующие буквы).

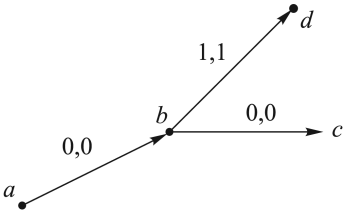

Рис. 1. Путь $a \to b \to c$ — барьер

Для того чтобы сделать процесс угадывания наглядным, введем следующее понятие. Назовем *барьером* двухреберный путь $a \to b \to c$ (рис. 1), удовлетворяющий условию $g(a,b) = f(b,c)$. Если дано слово и соответствующий ему путь, то правильное угадывание буквы слова происходит в том и только том случае, когда путь прошел через барьер.

Орграф с метками $f : E \to \{0,1\}$ и $g : E \to \{0,1\}$ обозначим через $G = (Q, E, f, g)$. Если фиксирована вершина q_0, то запишем $G = (Q, E, f, g, q_0)$[1]. Напомним, что g-метки заданы, возможно, не на всех ребрах. Итак, автомат $G = (Q, E, f, g)$ служит и для задания языка (по данному множеству путей), и для угадывания этого языка.

Степень угадывания и интервал угадывания. Введем понятие, позволяющее оценить "качество" данного процесса угадывания. Оно было введено в работе [2]; воспроизведем его, используя приведенную выше терминологию (отметим, что она отличается от терминологии, принятой в работе [3]).

Возьмем язык L. Пусть дан некоторый процесс угадывания; таким образом, для любого слова $\alpha = \alpha(1)\alpha(2)\ldots$, взятого из процесса L, указана "угадывающая процедура" (2) и цепочка прогнозов (1).

Возьмем слово α и соответствующую ему цепочку прогнозов (1). Для каждого $n \geqslant 1$ обозначим через $r(n)$ число удачных угадываний на временно́м промежутке $1, \ldots, n$; другими словами, $r(n) = |\{k : \alpha(k+1) = \beta(k)\}, k \leqslant n|$. *Степень угадывания* слова определяется равенством $\underline{\lim}\, r(n)/n, n \to \infty$. Степень угадывания языка L — такое число σ, что любое слово из процесса L угадывается со степенью, не меньшей σ. Как следует из определения, если σ — степень угадывания, то таковым будет и любое число $\sigma_1 < \sigma$.

Язык назовем *частично угадываемым в широком смысле*, если существует процесс угадывания (2) с ненулевой степенью угадывания. Далее модифицируем эти понятия следующим образом (снова используем цепочку прогнозов (1)).

Скажем, что слово α частично угадывается с *интервалом угадывания* T, если на любом промежутке $i, i+1, \ldots, i+T$ с достаточно

[1] В другой терминологии имеем неинициальный и, соответственно, инициальный автомат с входной функцией f и выходной функцией g.

большим i имеется момент угадывания (т.е. найдется такое j, что $\alpha(j+1)=\beta(j)$). Скажем также, что язык L *угадывается с интервалом* T, если каждое слово языка угадывается с интервалом T.

Язык назовем *частично угадываемым в узком смысле*, если найдется процесс угадывания (2) с конечным интервалом угадывания. Степень и интервал угадывания для путей определяется так же, как и для языков.

Легко заметить, что из частичной угадываемости в узком смысле следует частичная угадываемость в широком смысле. Причем в качестве степени угадывания можно взять число $1/T$ (а также любое меньшее число), где T — интервал угадывания.

Обратное утверждение неверно. Так, последовательность вида $(0^n(0\cup 1)^n)^\infty$ будет угадана в широком смысле автоматом, выдающим константу 0 (со степенью 0,5), но для любого угадывающего устройства в ней может встречаться произвольное число неугаданных подряд символов. Таким образом, возникает задача исследования угадываемости (в широком или узком смысле) языков и множеств путей. Далее наметим класс языков и множеств путей, для которых рассмотрим эту задачу.

План дальнейшего исследования задачи угадывания. Вернемся к леммам 1 и 2. Согласно этим леммам, намечается следующий естественный класс множеств путей и языков, для которых можно рассмотреть задачи угадывания:

1) множества путей вида $R(Q)$ и языки $L(q_0,Q)$ для сильно связного орграфа Q;
2) множества путей вида $R(Q,Q')$ и языки $L(q_0,Q,Q')$, где Q' — сильно связное множество вершин орграфа $[Q]$ (напомним, что $[Q]$ — краткое обозначение орграфа с множеством вершин Q);
3) объединения множеств путей и языков, указанных в предыдущем пункте.

Перечисленные задачи изучим последовательно.

Задача угадывания для множеств путей в сильно связных графах. *Множества $R(Q)$*. Рассмотрим сильно связный граф $[Q]$. Напомним, что он назван совершенным, если из каждой вершины выходят два ребра (и несовершенным — в противном случае). Ясно, что для совершенного орграфа множество $R(Q)$ не допускает частичного угадывания. Для несовершенного орграфа — положение иное.

Теорема 1. *Пусть дан сильно связный несовершенный орграф* $[Q]$. *Тогда множество* $R(Q)$ *частично угадываемо*, *причем интервалом угадывания будет число* $n, n=|Q|$. *Это верно и для языка* $L(q_0,Q)$.

Необходимый автомат-угадчик будет построен в ходе доказательства, которое начнем с доказательства леммы.

Лемма 3. *Пусть дан сильно связный орграф. Тогда любой его сильно связный подграф* $[Q'], [Q'] \neq [Q]$ *несовершенен.*

◀ Допустим, что подграф $[Q']$ совершенен. Возьмем вершины $a \in [Q']$ и $b \notin [Q']$. Поскольку граф $[Q]$ сильно связный, существует путь $v_1 \to \ldots \to v_r, v_1 = a, v_r = b$. При $a \in [Q']$ и $b \notin [Q']$ найдется такое ребро $v_i \to v_{i+1}$, что $v_i \in [Q']$ и $v_{i+1} \notin [Q']$. Итак, из вершины v_i вышло ребро $v_i \to v_{i+1}$, не содержащееся в множестве Q'. Это противоречит тому, что подграф $[Q']$ совершенен. ▶

Продолжим доказательство теоремы 1. Цель — подобрать g-метки так, чтобы на каждом пути длиной, превосходящей число n, встретился барьер. Опишем необходимую конструкцию.

Построим цепочку орграфов $[Q_0] \supset [Q_1] \supset \cdots \supset [Q_m]$ и снабдим некоторые ребра g-меткой. Для этого примем, во-первых, $[Q] = [Q_0]$. Если уже построен граф $[Q_k]$, то для построения следующего графа $[Q_{k+1}]$ возьмем сначала любую сильно связную компоненту $[Q']$ графа $[Q_k]$. Согласно лемме 3, $[Q']$ — несовершенный граф, а потому содержит вершину b_k, из которой выходит только одно ребро $b_k \to c_k$, лежащее в подграфе $[Q']$; назовем такое ребро *одиночным*. На каждом ребре $d \to b_k$ исходного графа $[Q]$, входящем в вершину b_k, зададим g-метку вида $g(d, b_k) = f(b_k, c_k)$; получим барьеры $d \to b_k \to c_k$. Удалим из орграфа $[Q_k]$ вершину b_k, ребро $b_k \to c_k$ и все ребра, входящие в вершину b_k. Получим новый орграф; это и есть искомый граф $[Q_{k+1}]$. Работа заканчивается тогда, когда очередной граф не имеет циклов и, как следствие, не имеет нетривиальных связных компонент.

Каждый граф $[Q_{k+1}]$ получается из предыдущего графа $[Q_k]$ удалением одной вершины b_k (и ребер, содержащих вершину b_k). Получаем также g-метки на некоторых ребрах. На остальных ребрах ставим g-метки произвольно. В результате исходный размеченный орграф $([Q], f)$ превратился в размеченный орграф $G = ([Q], f, g)$. Отметим, что граф, полученный после удаления указанных ребер, далее не используется; он необходим лишь как средство построения набора g-меток.

Для дальнейшего исследования введем следующие понятия: 1) *петля* — путь $q_1 \to \ldots \to q_s$, где $q_s = q_1$, а вершины $q_1, \ldots, q_{s-1}$ попарно различны (рис. 2, *а*); 2) *расширенная петля* — путь $q_0 \to \ldots \to q_s$, что $q_0 \neq q_1$ и путь $q_1 \to \ldots \to q_s$ есть петля (рис. 2, *б*).

Лемма 4. *Пусть в графе* $[Q]$ *дан путь и его отрезок, являющийся расширенной петлей. Тогда этот отрезок содержит барьер.*

◀ Возьмем расширенную петлю $q_0 \to \ldots \to q_s$ и соответствующую петлю $q_1 \to \ldots \to q_s$, где $q_s = q_1$. Найдется такое $r \geq 0$, что эта петля лежит в графе $[Q_r]$ и не лежит в графе $[Q_{r+1}]$ (так как построенная цепочка убывает начиная с исходного графа и заканчивая графом

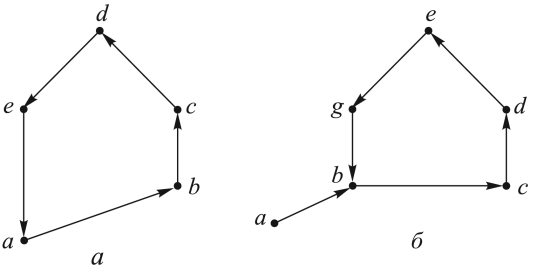

Рис. 2. Петля (*а*) и расширенная петля (*б*)

без циклов). Поскольку граф $[Q_{r+1}]$ получен из графа $[Q_r]$ удалением ровно одной вершины b_r, эта вершина лежит на петле. Возможны два случая: 1) вершина b_r совпадает с вершиной q_1; 2) она совпадает с некоторой вершиной $q_j, 1 < j < s$. В обоих случаях $b_r \to c_r$ — одиночное ребро в графе $[Q_r]$. Поэтому, и по построению функции g, в первом случае имеем барьер $q_0 \to q_1 \to q_2$, где $q_1 = b_r, q_2 = c_r$, во втором — барьер $q_{j-1} \to q_j \to q_{j+1}$, где $q_{j+1} = c_r$. ►

Для следующей леммы используем введенное обозначение $n = |Q|$.

Лемма 5. *Любой отрезок пути длиной, не меньшей* $n + 1$, *не являющийся начальным*, *содержит расширенную петлю.*

◄ Любой путь $q_1 \to \ldots \to q_m$ длиной, не меньшей n, содержит совпадающие вершины и, следовательно, петлю. Пусть эта петля имеет вид $q_i \to \ldots \to q_s, q_i = q_s$. Поскольку рассматриваемый отрезок не начало пути, на этом пути имеется вершина q_{i-1}. Получим расширенную петлю $q_{i-1} \to q_i \to \ldots \to q_s$. ►

Из лемм 4 и 5 получаем, что любой не начальный отрезок пути длиной, не меньшей $n+1$, содержит барьер. Отсюда следует теорема 1.

Частичное угадывание множеств $R(q_0, Q, Q')$ и языков $L(q_0, Q, Q')$. Вернемся к множествам $R(Q, Q')$, выделяемым условием $\underline{\lim}\, \alpha = Q'$. Известно, что эти множества $R(Q, Q')$ непусты тогда и только тогда, когда подграф $[Q']$ сильно связный. Из теоремы 1 вытекает следствие.

Следствие 1. *Множество* $R(Q, Q')$ *частично угадываемо тогда и только тогда*, *когда* $[Q']$ *— несовершенный подграф.*

Отсюда немедленно получаем соответствующее утверждение и для языков $L(q_0, Q, Q')$. Несколько модифицируем процедуру угадывания множества посредством g-меток в орграфе $G = (Q, E, f, g)$. Рассмотрим сильно связное множество $Q' \subset Q$ и соответствующее $R(Q')$. Пусть $[Q']$ — несовершенный подграф. Тем самым множество $R(Q, Q')$ угадываемо в узком смысле с некоторым интервалом угадывания. Условимся задавать g-метки, определяющие угадывание, только на ребрах, входящих в подграф $[Q']$. Если путь произволен, то процедура угадывания для него будет носить “прерывистый” характер: если

ребро $a_i \to a_{i+1}$ содержится в подграфе $[Q']$, то на этом ребре имеем прогноз (удачный, либо неудачный) относительно следующей вершины. Если это ребро не входит в подграф $[Q']$, то прогноз отсутствует. В частности, если указанный путь взят из $R(Q, Q')$, то начиная с некоторого момента прогнозы (возможно, неудачные) будут происходить уже без прерываний.

Выше речь шла о путях; изложенное можно перенести на порожденные ими языки.

Частичное угадывание объединения языков. Рассмотрим языки $L_s = L(q_s^0, Q_s, Q'_s)$, $s = 1, \ldots, m$, описанные выше. Пусть каждый язык $L_s = L(q_s^0, Q_s, Q'_s)$ частично угадываем (с некоторым интервалом угадывания T_s); таким образом, все орграфы $[Q'_s]$ несовершенны. Предположим также, что на каждом подграфе $[Q'_s]$ построены g_s-метки, задающие угадывание. Следовательно, имеем автоматы $G_s = (Q_s, E_s, f_s, g_s, q_{s0})$. Рассмотрим язык $L = \cup_{s=1}^m L(q_s^0, Q_i, Q'_i)$. Цель — построить алгоритм, позволяющий угадать этот язык (и оценить интервал угадывания).

Зафиксируем слово $\alpha = \alpha(1)\alpha(2)\ldots$ и постараемся построить для него цепочку прогнозов с положительным интервалом угадывания. Каждый автомат G_s с помощью своих f_s- и g_s-меток дает процесс угадывания (цепочка (2)), а также цепочку прогнозов (цепочка (1)), которая в рассматриваемом случае принимает вид

$$\beta_s(1), \beta_s(2), \ldots \tag{3}$$

Трудность в использовании этих прогнозов заключается в следующем. В каждый момент времени угадывающее лицо имеет в своем распоряжении несколько начальных букв угадываемого слова; однако эти начальные буквы слова не позволяют судить о том, какому из языков L_s принадлежит указанное слово. Поэтому процедура угадывания слова окажется не совсем простой.

Идея этой процедуры следующая. На каждый автомат подаем указанное слово $\alpha = \alpha(1)\alpha(2)\ldots$ Каждый автомат дает цепочку прогнозов относительно букв слова (напомним, что эта работа может иметь "прерывистый" характер, т.е. в некоторые моменты угадывание может и не происходить). Возьмем достаточно большое число R и из всех прогнозов, выданных автоматами в момент k, возьмем прогноз, принадлежащий автомату, который проработал (угадывал, возможно, не всегда успешно) без пропусков в предыдущие R моментов. (Если имеется несколько таких автоматов, то возьмем из них автомат с наименьшим номером.) Это и есть искомый прогноз относительно следующей буквы слова α.

Опишем намеченную процедуру подробно. Для каждого автомата G_s указанное слово $\alpha = \alpha(1)\alpha(2)\ldots$ дает путь $a_s(1) \to a_s(2) \to \ldots$ Обозначим через D_s множество натуральных чисел i, для которых ребро $a_s(i) \to a_s(i+1)$ лежит в множестве Q'_s. Известно, если $i \in D_s$, то в момент i автомат G_s выдает прогноз $\beta_s(i)$; если $i \notin D_s$, то в момент i автомат не дает никакого прогноза. Введем функцию $k \mapsto \Phi(k), k = 1, 2, \ldots$, по следующему правилу: для каждого k рассмотрим такие s, для которых числа $k, k-1, \ldots, k-R$ входят в множество D_s. Другими словами, выбираются те автоматы, которые давали прогноз от момента $k-R$ до момента k. Наименьший из таких номеров s обозначим через $\Phi(k)$. Итак, равенство $\Phi(k) = i$ означает следующее: во-первых, множество чисел $\{k-R, \ldots, k\}$ содержится в множестве D_i, во-вторых, указанное множество чисел не содержится в множестве $D_{i'}$ при $i' < i$. Функция Φ построена. Отметим ее свойство, которое следует из ее определения и будет использовано далее: если для некоторого i числа $\{k-R, \ldots, k\}$ входят в множество D_i, то $\Phi(k) \leq i$. Наконец, взяв автомат с найденным номером $i = \Phi(k)$, возьмем выданный им (в момент k) прогноз $\beta_i(k)$. В результате для угадывания слова $\alpha = \alpha(1)\alpha(2)\ldots$ получили цепочку прогнозов

$$\beta(k) = g_i(k),\ k = 1, 2, \ldots,\ \ i = \Phi(k). \tag{4}$$

В цепочке (4) возможны пропуски (отмечаем их черточками), так как для некоторых k величина $\Phi(k)$ может быть не определена. Образно говоря, в каждый момент времени k каждый m-й автомат пытается угадать символ $\alpha(k+1)$ нашего слова; но из всех прогнозов отбирается прогоз, выданный автоматом с наименьшим номером, который угадывал (возможно, не всегда удачно) в моменты времени от $k-R$ до k (без прерываний).

Теорема 2. *1. Функция $k \to \Phi(k)$ определена при $k > k_0$, где k_0 — постоянная, зависящая от α. 2. Пусть число R выбрано так, что $R > 4mT_i$ для всех $i = 1, \ldots, m$. Тогда последовательность (4) дает процесс частичного угадывания с интервалом угадывания R.*

◀ Докажем первое утверждение теоремы. По условию $\alpha \in L_s = L(q_s^0, Q_s, Q'_s)$ для некоторого s. Поскольку $\overline{\lim}\alpha = Q'_s$, соответствующий путь в орграфе G_s полностью лежит в Q'_s, начиная с некоторого момента k_0. Следовательно, числа $k, k-1, \ldots, k-R$ принадлежат множеству Q'_s для всех $k > k_0$, если число k_0 достаточно велико. Поэтому функция определена при $k > k_0$.

Доказательство второго утверждения теоремы 2 требует некоторых приготовлений. Число $k > k_0$ назовем точкой скачка функции Φ при $\Phi(k) \neq \Phi(k+1)$.

Лемма 6. *В любом отрезке длиной R имеется не более $2m$ точек скачка.*

◄ 1. Зафиксируем отрезок длиной R и рассмотрим точки из него, в которых выполнено неравенство $\Phi(k) < \Phi(k+1)$. Докажем, что число таких точек не превосходит m. Допустим противное. Поскольку функция $\Phi(k)$ принимает не более чем m значений, а число скачков превосходит m, найдутся такие k, l, что $\Phi(k) < \Phi(k+1), \Phi(l) < \Phi(l+1), k < l, \Phi(k) = \Phi(l) = i$ для некоторого i. Из определения функции Φ следует, что $[k-R, k] \subset D_i$ и $[l-R, l] \subset D_i$. Однако $l - R \leqslant k$, поэтому $[k-R, l] \subset D_i$. В частности, $[k+1-R, k+1] \subset D_i$. В соответствии с определением функции Φ, $\Phi(k+1) \leqslant i$, т.е. $\Phi(k+1) \leqslant \Phi(k)$. Противоречие.

2. Рассмотрим точки скачка, в которых выполнено неравенство $\Phi(k) > \Phi(k+1)$. Докажем, что число таких точек не превосходит m. Здесь рассуждения аналогичны предыдущим; тем не менее проведем их. Допустим противное и каждому скачку поставим в соответствие пару $(\Phi(k), \Phi(k+1))$. Тогда найдутся такие k, l, что $\Phi(k) > \Phi(k+1), \Phi(l) > \Phi(l+1), k < l, \Phi(k+1) = \Phi(l+1) = i$ для некоторого i. Из определения функции Φ следует, что $[k+1-R, k+1] \subset D_i$ и $[l+1-R, l+1] \subset D_i$. Однако $l+1-R \leq k+1$, поэтому $[k+1-R, l+1] \subset D_i$. В частности, $[l-R, l] \subset D_i$. Согласно определению функции Φ, $\Phi(l) \leqslant i$, т.е. $\Phi(l) \leqslant \Phi(l+1)$. Противоречие.

Учитывая пп. 1 и 2 получаем утверждение леммы 6. ►

Лемма 7. *В любой цепочке* $\{s, s+1, \ldots, s+R\}$ *существует отрезок длиной* $[R/(2m)]$*, на котором функция* Φ *постоянна.*

Доказательство леммы 7 следует из леммы 6.

Вернемся к доказательству второго утверждения теоремы 2.

Возьмем отрезок натуральных чисел длиной R, в нем отрезок длиной $[R/(2m)]$, указанный в лемме 7; пусть этот отрезок есть $\{l, l+1, \ldots, l+r\}$, $r = [R/(2m)]$. На этом отрезке функция Φ принимает постоянное значение i. Поэтому цепочка прогнозов (4) принимает вид $\beta_i(l), \beta_i(l+1), \ldots, \beta_i(l+r)$, т.е. совпадает с цепочкой прогнозов, полученных автоматом G_i. Напомним, что $[R/(2m)] > T_i$. Следовательно, на указанном отрезке имеется момент правильного угадывания для автомата G_i. Теорема 2 доказана. ►

Следствие 2. *Язык вида* $L = \cup_i^m L(q_0^i, Q_i, Q_i')$ *является частично угадываемым тогда и только тогда, когда каждый подграф* $[Q_i']$ *несовершенен, причем угадывание происходит с интервалом* $\max\{|Q_i'|, i = 1, \ldots, m\}$.

Заключение. Как было отмечено выше, изучение частичной угадываемости языков было начато в работах [2, 3]. Теорема 1 (критерий частичной угадываемости языка $L(q_0, Q, Q')$) была доказана в работе [3]. При этом и терминология, и методы, принятые в указанной работе, существенно отличались от тех, какие были использованы в

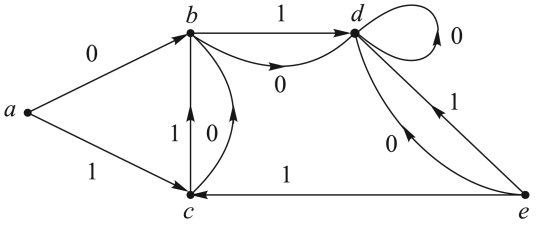

Рис. 3. Автомат, представляющий с помощью семейства множеств состояний $\{\{d\}\{d,e\}\}$ язык

настоящей работе. Далее в работе [3] доказано, что объединение языков вида $L(q_0, Q, Q')$ являтся частично угадываемым в широком смысле, когда каждый язык частично угадываем в широком смысле. Верно и обратное.

Для указанного объединения языков в работе [3] был использован термин "язык, представимый с помощью семейства множеств состояний" и объяснялась их роль в проблеме угадывания. Автомат, представляющий с помощью семейства множеств состояний $\{\{d\}\{d,e\}\}$ язык, который состоит из всех слов, где после некоторого момента не встречается две единицы подряд, приведен на рис. 3.

Общерегулярные языки были исследованы в работах [2, 3], в которых показано, что вопрос об их угадывании сводится к частичному угадыванию языков вида $L(q_0, Q, Q')$. Угадывание выполнялось с помощью конечных автоматов.

В настоящей статье изложение, приведенное в упомянутых работах, модифицировали, в частности, указали простые алгоритмы угадывания.

Завершая сопоставление работы [3] с настоящей, отметим существенный инструмент, использованный в работе [3]. Язык вида $L(q_0, Q, Q')$ — частично угадываемый тогда и только тогда, когда найдется слово, не содержащееся ни в одном слове этого языка. Этот критерий наличия "запрещенного" слова представляет самостоятельный интерес, но в данной работе не используется.

ЛИТЕРАТУРА

1. *Верeникин А.Г.*, *Гасанов Э.Э.* Об автоматной детерминизации множеств сверхслов // Дискретная математика. 2006. Т. 18. № 2. С. 84–97.
2. *Мастихина А.А.* О частичном угадывании сверхслов // Интеллектуальные системы 2007. Т. 11. Вып. 1–4. С. 609–619.
3. *Мастихина А.А.* Критерий частичного предвосхищения общерегулярных сверхсобытий // Дискретная математика. 2011. Т. 23. № 4. С. 103–114.
4. *Трахтенброт Б.А.*, *Бардзин Я.М.* Конечные автоматы (поведение и синтез). М.: Наука, 1970.
5. *Мастихина А.А.* Частичное угадывание сверхсобытий, порожденных простыми $LL(1)$-грамматиками // Интеллектуальные системы. 2011. Т. 15. С. 507–532.

REFERENCES

[1] Verenikin A.G., Gasanov E.E. On the automaton determinization of sets of superworks. *Discrete Mathematics and Applications*, 2006, vol. 16, iss. 3, pp. 229–243.

[2] Mastikhina A.A. On the partial guessing of superwords. *Intellekt. Sist.* [Intelligent Systems], 2007, vol. 11, iss. 1–4, pp. 609–619 (in Russ.).

[3] Mastikhina A.A. A criterion for a partial prediction of general regular superevents. *Discrete Mathematics and Applications*, 2011, vol. 21, iss. 5–6, pp. 727–739.

[4] Trakhtenbrot B.A., Bardzin Ya.M. Konechnye avtomaty: povedenie i sintez [Finite automata: Behavior and synthesis]. Moscow, Nauka Publ., 1970.

[5] Mastikhina A.A. Partial Guessing Superevents Generated by the Simple $LL(1)$ Grammars. *Intellekt. Sist.* [Intelligent Systems], 2011, vol. 15, pp. 507–532 (in Russ.).

Статья поступила в редакцию 29.06.2015

Исмагилов Раис Сальманович — д-р физ.-мат. наук, профессор кафедры "Высшая математика" МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Ismagilov R.S. — Dr. Sci. (Phys.-Math.), Professor of Higher Mathematics Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Мастихина Анна Антоновна — канд. физ.-мат. наук, доцент кафедры "Высшая математика" МГТУ им. Н.Э. Баумана (Российская Федерация, 105005, Москва, 2-я Бауманская ул., д. 5).

Mastikhina A.A. — Cand. Sci. (Phys.-Math.), Assoc. Professor of Higher Mathematics Department, Bauman Moscow State Technical University (2-ya Baumanskaya ul. 5, Moscow, 105005 Russian Federation).

Просьба ссылаться на эту статью следующим образом:

Исмагилов Р.С., Мастихина А.А. К вопросу частичного угадывания формальных языков // Вестник МГТУ им. Н.Э. Баумана. Сер. Естественные науки. 2016. № 2. С. 3–15. DOI: 10.18698/1812-3368-2016-2-3-15

Please cite this article in English as:

Ismagilov R.S., Mastikhina A.A. To the problem of partial guessing of formal languages. *Vestn. Mosk. Gos. Tekh. Univ. im. N.E. Baumana, Estestv. Nauki* [Herald of the Bauman Moscow State Tech. Univ., Nat. Sci.], 2016, no. 2, pp. 3–15.
DOI: 10.18698/1812-3368-2016-2-3-15