

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ  
М.В.ЛОМОНОСОВА

На правах рукописи  
УДК 519.71

Мастихина Анна Антоновна

ЧАСТИЧНОЕ ПРЕДВОСХИЩЕНИЕ СВЕРХСОБЫТИЙ  
АВТОМАТАМИ.

01.01.09 — дискретная математика и математическая кибернетика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени  
кандидата физико-математических наук

МОСКВА — 2012

Работа выполнена на кафедре математической теории интеллектуальных систем  
Механико-математического факультета Московского государственного университета  
имени М.В.Ломоносова.

<b>Научный руководитель</b>	доктор физико-математических наук, профессор Гасанов Эльяр Эльдарович
<b>Официальные оппоненты</b>	Подловченко Римма Ивановна, доктор физико-математических наук, профессор, Научно-исследовательский вычислительный центр Московского государственного университета имени М.В.Ломоносова, ведущий научный сотрудник  Лялин Илья Викторович, кандидат физико-математических наук, Элайн Технолоджи, старший разработчик
<b>Ведущая организация</b>	Российский государственный гуманитарный университет (РГГУ)

Защита диссертации состоится 8 июня 2012 г. в 16 ч. 45 мин. на заседании диссертационного совета Д.501.001.84 при Московском государственном университете имени М.В.Ломоносова по адресу: Российская Федерация, 119991, Москва, ГСП-1, Ленинские горы, д.1, МГУ, Механико-математический факультет, аудитория 14-08.

С диссертацией можно ознакомиться в библиотеке Механико-математического факультета МГУ (Главное здание, 14 этаж).

Автореферат разослан 27 апреля 2012 г.

Ученый секретарь диссертационного  
совета Д.501.001.84 при МГУ,  
д.ф.-м.н., профессор

Иванов А.О.

# ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

## Актуальность темы.

Понятие автомата является одним из центральных понятий в математической кибернетике. Задачи анализа поведения автоматов и синтеза автоматов с заданным поведением являются одними из ключевых задач теории автоматов. В диссертации исследуется вопрос о возможности использования автомата для предсказания своей входной последовательности и рассматривается задача синтеза автоматов, наилучшим образом предсказывающих некоторые наперед заданные множества входных последовательностей. Эта задача является актуальной для теории автоматов и представляет как теоретическую, так и практическую ценность, поскольку может быть использована во многих прогнозирующих системах и в технических устройствах, таких, например, как биржевые системы, оптимизирующие компиляторы, устройства обработки видео-изображений, фильтры шумоподавления и т.п.

Понятие автомата с предвосхищением появилось еще в начале развития теории автоматов<sup>1</sup>. Такой автомат использует в своей работе не только данные, уже поступившие на вход, но и некоторые значения, которые должны поступить в будущем. Используя эти значения, можно существенно ускорить работу устройства, а в случае, если предполагаемый символ не совпал с поступившим на вход на самом деле, происходит откат алгоритма, и выигрыша во времени не случается.

Соответственно, большое значение имеет выбор предполагаемого следующего символа. В качестве еще не поступивших данных часто используется наиболее вероятное значение, то есть для последовательности, символы которой нужно предсказать, известны вероятности встретить ту или иную комбинацию, либо эта вероятность вычисляется на основе поступившей части последовательности<sup>2</sup>.

Также можно использовать для предсказания некоторую информацию о строении последовательности. Например, если она порождена некоторой

<sup>1</sup>Трахтенброт Б.А., Бардзин Я.М. Конечные автоматы (поведение и синтез) // М.: Наука, 1970

<sup>2</sup>Hutter M. Optimality of Universal Bayesian Sequence Prediction for General Loss and Alphabet// Journal of Machine Learning Research N.4. — 2003. — Pp. 971-1000.

формальной грамматикой. Наиболее популярны регулярные<sup>3</sup> и контекстно-свободные грамматики<sup>4</sup>, они используются в языках программирования и относительно просты для разбора.

В статье А.Г.Вереникина и Э.Э.Гасанова<sup>5</sup> были введены угадывающие автоматы — конечные автоматы, угадывающие сверхслово или множество сверхслов. Это значило, что через некоторое конечное время после начала подачи сверхслова автомат начинает угадывать каждый следующий символ, то есть на выходе в момент времени  $t$  выдавать элемент входной последовательности с номером  $t + 1$ . Было показано, что угадываемы только множества периодических сверхслов.

В статье автора диссертации<sup>6</sup> было введено понятие частичного угадывания, которое имеет место в случае, если автомат угадывает следующий символ не обязательно в каждый момент времени, но достаточно часто.

Вводится степень угадывания как предел отношения числа верно предсказанных символов за время  $t$  от начала слова к длине префикса  $t$  при  $t$ , устремленном к бесконечности.

В зависимости от того, брать ли нижний или верхний предел в случае, если они не равны, меняется смысл задачи. В первом случае угадывание должно происходить регулярно, вне зависимости от того, когда измерять долю угаданных символов. Случай же верхнего предела иллюстрирует ситуацию, когда у угадывающего есть возможность самостоятельно выбирать, когда подсчитывать эту долю. Частичная угадываемость исследуется и для верхней, и для нижней степени.

Степень угадывания множества сверхслов (сверхсобытия) вводится как точная нижняя грань степени по всем сверхсловам этого множества. Таким образом, удачным считается случай, если на вход некоторой машины подается неизвестная последовательность из нулей и единиц из известного множества, и машина гарантированно предсказывает некоторую ненулевую долю символов.

---

<sup>3</sup>Maurel D., Pevedic B. The syntactic prediction with token automata: application to HandiAS system // Theoretical Computer Science. — 2001. — Vol. 267, Issue 1-2. — Pp.121-129.

<sup>4</sup>Wiles J., Elman J. L. Learning to count without a counter: A case study of dynamics and activation landscapes in recurrent networks. // Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society. - MIT Press. - 1995. - Pp. 482Ц487.

<sup>5</sup>Вереникин А.Г., Гасанов Э.Э. Об автоматной детерминизации множеств сверхслов // Дискретная математика. — 2006. — Т.18, №2. — С. 84–97.

<sup>6</sup>Мастихина А.А. О частичном угадывании сверхслов // Интеллектуальные системы. — 2007. — Т.11, вып.1-4 — С.609-619.

**Цель работы.** Цель работы состоит в нахождении критериев, позволяющих определить, является ли некоторое сверхсобытие частично угадываемым. В случае положительного ответа необходимо построить автомат, угадывающий данное сверхсобытие с как можно большей степенью.

**Методы исследования.** В работе используются методы дискретной математики, а именно, теории автоматов и теории графов, также математического анализа.

**Научная новизна.**

Диссертация имеет теоретический характер. Результаты получены лично автором, являются новыми и состоят в следующем.

1. Получен критерий частичной угадываемости общерегулярных множеств. В случае, если множество угадываемо, строится угадывающий конечный автомат, причем он отличается от принимающего только выходной функцией. Также предложен способ построения угадывающего автомата, основанный на информации о словах, не встречающихся во множестве последовательностей. Получен способ нахождения для общерегулярного множества наибольшей возможной степени и построения автомата, угадывающего его с этой степенью.

2. Получен критерий частичной угадываемости сверхитераций языков, порожденных простыми  $LL(1)$ -грамматиками. Для частично угадываемых сверхсобытий строится угадывающий алгоритм, реализуемый автоматом с магазинной памятью. Частичная угадываемость рассматривается в смысле верхнего и нижнего пределов доли верно предсказанных символов. Критерии и алгоритмы для этих двух случаев различаются.

3. Для сверхитераций детерминированных контекстно-свободных языков, сохраняющих детерминированность при итерации, получен критерий частичной угадываемости в смысле нижней степени.

4. Получены оценки степени угадывания, причем они достижимы, то есть для множеств с наименьшей возможной долей предложенный алгоритм угадывает именно столько.

Идея частичного угадывания состоит в том, что в случае, когда оно вообще возможно, последовательности должны обладать некоторыми ограничениями, которые нарушаются, только если последовательность оказы-

вается не из данного множества. Поэтому в определенные моменты существует лишь один вариант следующего символа, и происходит угадывание. В остальные моменты выходная функция предсказывает такие символы, которые отдалают момент гарантированного угадывания, то есть если угадывания не происходит, то автомат приближается к состоянию, в котором угадывание гарантировано. Поэтому угадывание происходит достаточно часто.

**Практическая ценность.** В работе понятие угадывания заданной последовательности заданным алгоритмом вводится как предел доли угаданных символов. Частичное угадывание множества последовательностей имеет место, если алгоритм угадывает часть символов каждой последовательности, то есть даже в худшем случае есть некоторый выигрыш.

Рассмотренные машины, реализующие алгоритмы угадывания — детерминированные автоматы — представляются естественными для этой цели.

Множества, все последовательности которых нельзя угадать никаким детерминированным автоматом, случайны в том смысле, что в каждый момент следующий символ может оказаться любым. Полученные критерии иллюстрируют, как отличается это свойство для множеств, порожденных разными формальными грамматиками.

Допустимы обобщения, такие как подсчет средней доли угадывания во всем множестве, а также комбинации с вероятностным подходом.

Результаты могут быть использованы в программировании при создании оптимизирующих компиляторов. Регулярные и  $LL(1)$ -грамматики широко распространены в этой области, так как они обладают широкими выразительными возможностями, но при этом достаточно просты для распознавания. Угадывающие автоматы могут иметь широкое применение при синтезе чипов для обработки интенсивных потоков данных, поскольку в этих устройствах без предсказания поступающих последовательностей невозможно добиться нужной производительности.

Для рассмотренных множеств критерии разграничивают случаи, когда имеет смысл предсказывать, основываясь на поступившей части информационного потока и общей структуре множества, и когда никакой алгоритм не сможет дать результата, лучшего, чем подбрасывание монеты.

Алгоритмы, предложенные для частичного угадывания, удобны, во-первых, тем, что практически не требуют специальных построений, так как основываются на устройстве, принимающем данное множество последовательностей. Во-вторых, среди некоторых множеств они угадывают существенную долю символов, иногда максимально возможную.

**Апробация работы.** Результаты данной работы докладывались и публиковались в тезисах следующих конференций.

Конференция студентов, аспирантов и молодых ученых мех-мат. ф-та МГУ (2008, Москва).

Международная конференция "Современные проблемы математики, механики и их приложений", посвященная 70-летию академика В.А.Садовниченко (2009, Москва).

Международная конференция студентов, аспирантов и молодых ученых "Ломоносов"(2009, 2010, 2011, 2012, Москва). Доклады на конференциях "Ломоносов-2010", "Ломоносов-2011" и "Ломоносов-2012" были отмечены грамотами за лучший доклад.

Международный семинар "Дискретная математика-2010" (2010, Москва).

54-ая научная конференция МФТИ "Проблемы фундаментальных и прикладных естественных и технических наук в современном информационном обществе"(2011, Москва-Долгопрудный-Жуковский).

X Международная конференция "Интеллектуальные системы и компьютерные науки" (Москва, 2011).

Также результаты докладывались на семинарах механико-математического факультета МГУ им. М.В.Ломоносова: на семинаре "Теория автоматов" под руководством академика, проф., д.ф-м.н. В.Б. Кудрявцева (2011 г.), на семинаре "Вопросы сложности алгоритмов поиска" под руководством проф., д.ф-м.н. Э.Э.Гасанова (2006-2011 гг.), на семинаре МИАН и ВЦ по теории сложности под руководством член-корр., д.ф-м.н. А.А.Разборова (2011 г.), на семинарах факультета ВМиК МГУ им. М.В.Ломоносова: на семинаре "Дискретная математика и математическая кибернетика" под руководством проф., д.ф-м.н. В.Б.Алексеева, проф., д.ф-м.н. А.А.Сапоженко, проф., д.ф-м.н. С.А.Ложкина (2011 г.), на

семинаре "Теоретические проблемы программирования" под руководством проф., д.ф-м.н. Р.И.Подловченко, д.ф-м.н. В.А.Захарова (2012 г.).

**Публикации.** По теме диссертации опубликовано 10 работ, список которых приведен в конце автореферата [1-10].

**Структура и объем работы.** Диссертация состоит из введения и 4 глав. Объем диссертации 95 страниц. Список литературы содержит 25 наименований.

## ОСНОВНЫЕ ПОНЯТИЯ

Будем рассматривать класс детерминированных, но, возможно, бесконечных автоматов  $\mathfrak{M}$ .

Через  $y_\alpha^{\mathfrak{A}}$  обозначим выходное сверхслово некоторого автомата  $\mathfrak{A}$  при подаче на его вход сверхслова  $\alpha$ . Детерминированность автомата означает, что после подачи на его вход  $t$  первых символов сверхслова  $\alpha$  он однозначно выдает некоторый выходной символ  $y_\alpha^{\mathfrak{A}}(t)$ .

Если  $\alpha \in \{0, 1\}^\infty$ , то обозначим

$$d^{\mathfrak{A}}(\alpha, t) = \frac{1}{t} \sum_{i=1}^t (1 - |y_\alpha^{\mathfrak{A}}(i) - \alpha(i+1)|).$$

Автомат  $\mathfrak{A}$  *угадывает* сверхслово  $\alpha \in \{0, 1\}^\infty$  с (нижней) степенью  $\sigma \in [0, 1]$ , если

$$c^{\mathfrak{A}}(\alpha) = \underline{\lim}_{t \rightarrow \infty} d^{\mathfrak{A}}(\alpha, t) = \sigma.$$

$\mathfrak{A}$  *угадывает* сверхслово  $\alpha \in \{0, 1\}^\infty$  с верхней степенью  $\sigma \in [0, 1]$ , если

$$\bar{c}^{\mathfrak{A}}(\alpha) = \overline{\lim}_{t \rightarrow \infty} d^{\mathfrak{A}}(\alpha, t) = \sigma.$$

Величины  $c^{\mathfrak{A}}(\alpha)$  и  $\bar{c}^{\mathfrak{A}}(\alpha)$  будем называть соответственно *степенью угадывания (предвосхищения)* и *верхней степенью угадывания* сверхслова  $\alpha$  автоматом  $\mathfrak{A}$ .

Автомат  $\mathfrak{A}$  частично угадывает множество сверхслов  $A$  (частично угадывает в смысле верхнего предела), если  $\exists \sigma > 0$ , такое что  $\forall \alpha \in A$  выполнено  $c^{\mathfrak{A}}(\alpha) > \sigma > 0$  ( $\bar{c}^{\mathfrak{A}}(\alpha) > \sigma > 0$ ). Множество  $A$  частично угадываемо, если найдется частично угадывающее его  $\mathfrak{A} \in \mathfrak{M}$ .



Степень угадывания множества  $A$

$$c^{\mathfrak{A}}(A) = \inf_{\alpha \in A} c^{\mathfrak{A}}(\alpha).$$

Также будут рассматриваться автоматы без выходов  $\mathfrak{A} = (A, Q, \varphi, q_0)$ .

Автомат  $\mathfrak{A} = (A, Q, \varphi, q_0)$  допускает сверхсобытие  $R$  с помощью выделенного множества семейств состояний  $F \subseteq 2^Q$ , если  $\alpha \in R$  тогда и только тогда, когда множество состояний, проходимых автоматом бесконечное число раз при подаче на его вход сверхслова  $\alpha$ , есть элемент  $F$ .

В любом автомате можно выделить сильно связанные компоненты  $Q_1, \dots, Q_k$ , то есть совокупности состояний, достижимых друг из друга:  $\forall q', q'' \in Q_i, i = 1, \dots, k, \exists \alpha \in \{0, 1\}^*$ , т.ч.  $\varphi(q', \alpha) = q''$ .

Назовем сильно связную компоненту замкнутой, если из нее не достижима никакая другая сильно связанная компонента.

Контекстно-свободная грамматика  $(N, T, P, S)$  (где  $N$  — множество нетерминальных символов (НС),  $T$  — множество терминальных символов,  $S \in N$  — аксиома,  $P$  — множество правил вывода) находится в нормальной форме Грейбах, если все правила имеют вид  $A \rightarrow aB_1 \dots B_k$ , где  $B_1, \dots, B_k \in N, a \in T$ , или  $A \rightarrow a$ , или  $S \rightarrow \lambda$ , где  $\lambda$  — символ для обозначения пустого слова. Причем, если грамматика простая  $LL(1)$ , для каждого нетерминального символа есть не более  $|T|$  правил (с разными терминальными символами). Любую грамматику можно привести к нормальной форме Грейбах (то есть найти грамматику в нормальной форме Грейбах, порождающую тот же язык)<sup>7</sup>. Причем простая  $LL(1)$ -грамматика при приведении к нормальной форме Грейбах остается простой  $LL(1)$ .

Символ  $A$  достижим из символа  $B$ , если найдутся такие  $\alpha \in T^*, \beta \in (T \cup N)^*$ , что  $B \Rightarrow^* \alpha A \beta$ .

Символ грамматики  $A$  называется достижимым, если найдутся такие  $\alpha \in T^*, \beta \in (T \cup N)^*$ , что  $S \Rightarrow^* \alpha A \beta$ . Иначе символ называется недостижимым.

Нетерминальный символ  $A$  называется тупиковым, если не найдется такого слова  $\beta \in T^*$ , что  $A \Rightarrow^* \beta$ .

<sup>7</sup>Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции: Пер. с англ. М.: Мир, 1978.

Далее будем рассматривать грамматики в нормальной форме Грейбах и считать, что недостижимые и тупиковые символы отсутствуют. Также будем считать, что аксиома не встречается в правой части правил. Будем тогда говорить, что грамматика находится в нормальной форме.

Автомат с магазинной памятью — это семерка  $(Q, \Sigma, \Gamma, Z, \Phi, q_0, F)$ , где  $Q$  — множество состояний,  $\Sigma$  — входной алфавит,  $\Gamma$  — магазинный алфавит,  $Z$  — символ для обозначения пустого магазина,  $q_0 \in Q$  — выделенное начальное состояние,  $F \subseteq Q$  множество заключительных состояний,  $\Phi$  — множество команд, каждая команда — отображение  $(Q \times (\Sigma \cup \lambda) \times \Gamma \rightarrow Q \times \Gamma^*)$ .

Конфигурацией будем называть пару  $(q, \xi)$ ,  $q \in Q, \xi \in \Gamma^*$ : текущее состояние и содержимое магазина.

Автомат работает потактово. Такт — работа автомата после поступления одного символа. Считаем, что  $\lambda$ -такт выполняется мгновенно.

Автомат называется *детерминированным*, если для каждой конфигурации  $(q, \xi)$  имеется либо не более одной команды вида  $(q, a, \xi]_1 \rightarrow r, \beta)$ ,  $q, r \in Q, \xi, \beta \in \Gamma^*$  для каждого  $a \in \Sigma$  и ни одной команды вида  $(q, \lambda, \xi]_1 \rightarrow r', \beta')$ ,  $r' \in Q, \beta' \in \Gamma^*$ , либо ни одной команды вида  $(q, a, \xi]_1 \rightarrow r, \beta)$  и не более одной вида  $(q, \lambda, \xi]_1 \rightarrow r', \beta')$ .

## СОДЕРЖАНИЕ РАБОТЫ

В первой главе рассматривается частичное угадывание в смысле нижнего предела конечными автоматами одного сверхслова. Получены оценки на число состояний автомата.

Пусть  $\mathfrak{B}_\sigma(\alpha)$  — множество всех автоматов, угадывающих сверхслово  $\alpha$  со степенью не меньшей  $\sigma$ .

Обозначим через  $\omega(\mathfrak{A})$  число состояний автомата  $\mathfrak{A}$  и введем

$$R_\sigma(\alpha) = \min_{\mathfrak{A} \in \mathfrak{B}_\sigma(\alpha)} \omega(\mathfrak{A})$$

— минимальное число состояний автомата, угадывающего сверхслово  $\alpha$  со степенью, не меньшей  $\sigma$ .

Пусть  $A(n)$  — множество сверхслов с периодом  $n$ .

**Теорема 1.** 1) Если  $\sigma = 1$ , то для любого  $n \in \mathbb{N}$  и любого  $\beta \in A(n)$  выполнено

$$\frac{m}{2} \leq R_\sigma(\beta) \leq m \leq n,$$

где  $m$  — длина минимального периода сверхслова  $\beta$ .

2) Если  $\sigma \leq \frac{1}{2}$ , то для любого  $n \in \mathbb{N}$  и любого  $\beta \in A(n)$  выполнено

$$R_\sigma(\beta) = 1.$$

3) Если  $\sigma > \frac{1}{2}$ , то для любого натурального  $n$  найдется такое периодическое сверхслово  $\gamma$ , что  $R_\sigma(\gamma) > n$ .

**Теорема 2.** Существует такое сверхслово, что ни один конечный автомат не угадывает его ни с какой степенью  $\sigma \in (0, 1]$ .

**Теорема 3.** 1) Для любого  $\sigma \in [0, \frac{1}{2}]$  существует сверхслово, которое угадывается любым конечным автоматом со степенью большей или равной  $\sigma$ .

2) Ни для какого  $\sigma \in (\frac{1}{2}, 1]$  не существует такого сверхслова, что все автоматы угадывают его со степенью большей или равной  $\sigma$ .

Во второй главе рассматривается частичное угадывание общерегулярных сверхсобытий конечными автоматами.

Получен критерий частичной угадываемости для общерегулярных сверхсобытий. Также получен алгоритм построения автомата, угадывающего сверхсобытие с наибольшей возможной степенью.

**Теорема 5.** Пусть общерегулярное сверхсобытие  $R^\infty$  представимо в конечном инициальном автомате  $\mathfrak{A} = (\{0, 1\}, Q, \varphi, q_0)$ ,  $|Q| = n$  с помощью некоторого  $F \subset 2^Q$ , тогда следующие утверждения эквивалентны.

1. Сверхсобытие  $R^\infty$ ,  $R \subseteq \{0, 1\}^*$ , не является частично угадываемым;
2.  $\forall \alpha \in \{0, 1\}^* \exists \beta \in \{0, 1\}^*, \gamma \in \{0, 1\}^\infty$ , что  $\beta\alpha\gamma \in R^\infty$ ;
3. Хотя бы для одной замкнутой сильно связной компоненты  $\tilde{Q}$  автомата  $\mathfrak{A}$  выполнено  $\tilde{Q} \in F$ .

**Теорема 7.** Общерегулярное сверхсобытие  $R'_1 R_1^\infty \cup R'_2 R_2^\infty \cup \dots \cup R'_k R_k^\infty$ , где  $R_1, R_2, \dots, R_k, R'_1, R'_2, \dots, R'_k$  — некоторые регулярные события, частично угадываемо тогда и только тогда, когда частично угадываемо каждое из сверхсобытий  $R_1^\infty, R_2^\infty, \dots, R_k^\infty$ .

**Теорема 8.** Пусть общерегулярное сверхсобытие  $L \subseteq \{0, 1\}^\infty$  представимо в конечном инициальном автомате  $\mathfrak{A} = (\{0, 1\}, Q, \varphi, q_0)$ ,  $|Q| = n$  с помощью некоторого  $F \subset 2^Q$ . Сверхсобытие  $L$  не является частично угадываемым тогда и только тогда, когда хотя бы для одной замкнутой сильно связной компоненты  $\tilde{Q}$  автомата  $\mathfrak{A}$  выполнено  $\tilde{Q} \in F$ .

Обозначим класс автоматов, отличающихся от  $\mathfrak{A}$  только добавлением функции выхода,  $\Phi(\mathfrak{A}) = \{(\{0, 1\}, Q, \{0, 1\}, \varphi, \psi, q_0), \psi : Q \rightarrow \{0, 1\}\}$ .

**Теорема 9.** Рассмотрим общерегулярное множество  $L$  и автомат  $\mathfrak{A} = (\{0, 1\}, Q, \varphi, q_0)$ , принимающий это множество с помощью  $F \subseteq 2^Q$ , при этом  $\forall \tilde{Q}_i, \tilde{Q}_j \in F, i \neq j$  выполнено  $\tilde{Q}_i \cap \tilde{Q}_j = \emptyset$ .

Тогда  $\forall g \in \mathfrak{M}$  выполнено  $c^g(L) \leq \max_{\mathfrak{A}' \in \Phi(\mathfrak{A})} c^{\mathfrak{A}'}(L)$ .

Как следствие, возникает следующая теорема.

**Теорема 10.** Существует алгоритм, определяющий, является ли сверхсобытие  $L$ , представимое в приведенном автомате с  $n$  состояниями, частично угадываемым, и его сложность линейно зависит от  $n$ . Также существует алгоритм построения минимального автомата с наибольшей возможной степенью предвосхищения в случае, если сверхсобытие частично угадываемо.

В третьей главе рассматривается частичное угадывание сверхитераций простых  $LL(1)$ -грамматик. Угадывание производится автоматами с магазинной памятью, отдельно рассмотрен случай нижней и верхней степеней. Найдены оценки для степеней угадывания и грамматики, для которых эти оценки достигаются.

**Теорема 11.** Множество сверхслов, образованное сверхитерацией языка  $L(G)$ , порожденного простой  $LL(1)$ -грамматикой  $G$  в нормальной форме, частично угадываемо в смысле нижней степени тогда и только тогда, когда хотя бы для одного нетерминального символа грамматики существует только одно правило вывода.

Обозначим за  $\tilde{G}(n, l)$  класс простых  $LL(1)$ -грамматик в нормальной форме с параметрами  $|N| = n$  и максимальной длиной команды (числом нетерминальных символов справа в команде)  $l$ .

За  $\tilde{G}^0(n, l)$  обозначим класс таких грамматик из  $\tilde{G}(n, l)$ , для которых  $L(G)^\infty$  частично угадываемо.

**Теорема 12.** Для произвольных  $n \geq 2, l \geq 1$  выполнено

$$\inf_{G \in \tilde{G}^0(n, L)} \sup_{\alpha \in \mathfrak{M}} c^\alpha(L(G)^\infty) = \frac{l-1}{2l^{n-1} - l^{n-2} - 1},$$

$$\inf_{G \in \tilde{G}^0(1, l)} \sup_{\alpha \in \mathfrak{M}} c^\alpha(L(G)^\infty) = 1.$$

Назовем правило грамматики *регулярным*, если оно имеет вид  $A \rightarrow aB$  (глубина стека соответствующего МП-автомата не увеличивается),  $A, B \in N$ . Назовем правило *стирающим*, если правило имеет вид  $A \rightarrow a$  (глубина стека уменьшается). Остальные правила назовем *удлинняющими*.

НС  $B$  *достижим по регулярным правилам по слову  $\alpha$  из НС  $A$* , если существуют регулярные правила

$$\begin{aligned} A &\rightarrow \alpha(1)A_1, \\ A_1 &\rightarrow \alpha(2)A_2, \\ &\dots \\ A_{|\alpha|} &\rightarrow \alpha(|\alpha|)B. \end{aligned}$$

Если существует такое слово  $\alpha$ , то  $B$  *достижим из  $A$  по регулярным правилам*.

Назовем *регулярной сильно связной компонентой  $\tilde{R}$*  множество НС, попарно достижимых друг из друга по регулярным правилам.

*Выходом* из регулярной сильно связной компоненты  $\tilde{R}$  назовем правило для  $A \in \tilde{R}$ , не являющееся регулярным или имеющее вид  $A \rightarrow aB$ , где  $B \notin \tilde{R}$ .

Опишем алгоритм  $\mathfrak{D}$ , разделяющий НС грамматики  $G = (N, \{0, 1\}, P, S)$  на классы следующим образом.

1-й шаг.

Рассмотрим регулярные сильно связные компоненты, все выходы из которых являются стирающими, занесем их в класс  $D_1$ .

Далее удалим из грамматики НС из класса  $D_1$ . То есть преобразуем грамматику  $G$  в грамматику  $G' = (N', \{0, 1\}, P', S)$ , где  $N' = N \setminus D_1$ , а  $P'$  получается из  $P$  удалением правил, начинающихся с элементов  $D_1$  и вычеркиванием элементов  $D_1$  из правых частей других содержащих их правил.

Например, если  $D_1 = \{A\}$ , то правило  $B \rightarrow aAC$  из  $P$  в  $P'$  будет иметь вид  $B \rightarrow aC$ .

Пусть сделано  $i$  шагов.

Опишем  $(i + 1)$ -й шаг.

В грамматике  $G^{(i)}$  рассмотрим сильно связанные регулярные компоненты, все выходы из которых являются стирающими. Все НС из таких компонент занесем в класс  $D_{i+1}$ .

Удалим из правил грамматики правила, начинающиеся с НС из класса  $D_{i+1}$ , а также вычеркнем элементы  $D_{i+1}$  из других правил. Получившуюся грамматику назовем  $G^{(i+1)}$ .

Алгоритм останавливается, когда некоторый класс  $D_k$  оказывается пуст или на  $k$ -ом шаге объединение классов  $\cup_{i=1}^k D_i$  есть множество всех НС  $N$ .

**Теорема 13.** *Множество сверхслов  $L(G)^\infty$ , где  $G = (N, \{0, 1\}, P, S)$  — простая  $LL(1)$ -грамматика в нормальной форме, частично угадываемым в смысле верхней степени тогда и только тогда, когда после работы алгоритма  $\mathfrak{D}$  выполнено  $\cup_{i=1}^k D_i \neq N$ , где  $k$  — номер шага, на котором алгоритм  $\mathfrak{D}$  остановился.*

Обозначим  $l_{min}(G)$  — минимальная длина правила грамматики  $G$ .

Обозначим  $r_c^i(G)$  наибольшую длину цепи регулярных правил в грамматике  $G^{(i)}$ , то есть наибольшее  $r$ , такое что существуют  $A_1 \rightarrow a_1 A_2, A_2 \rightarrow a_2 A_3, \dots, A_r \rightarrow a_r A_{r+1} \in P^{(i)}$ , причем  $A_1, \dots, A_{r+1}$  попарно различны.

$$R_c(G) = \max_{i=0, \dots, k} r_c^i(G).$$

Обозначим за  $\tilde{G}(l', l, R_c)$  класс простых  $LL(1)$ -грамматик  $G$  с параметрами  $l_{min}(G) = l', l_{max}(G) = l, R_c(G) = R_c$ .

За  $\tilde{G}^0(l', l, R_c)$  обозначим класс таких грамматик из  $\tilde{G}(l', l, R_c)$ , которые являются частично угадываемыми.

**Теорема 14.** *Для произвольных  $l \geq 1, R_c, l' \geq 1$  выполнено*

$$\inf_{G \in \tilde{G}(l, L, R_c)} \sup_{\mathfrak{A} \in \mathfrak{M}} \bar{c}^{\mathfrak{A}}((L(G))^\infty) = \min \left\{ \frac{l'}{l' + 1}, \frac{1}{R_c + 1}, \frac{l - 1}{l(R_c(l - 1) + l)^k - 1} \right\},$$

где  $k$  — номер шага, на котором алгоритм  $\mathfrak{D}$ , примененный к  $G$ , останавливается.

В четвертой главе рассмотрено частичное угадывание в смысле нижней степени сверхитераций детерминированных контекстно-свободных языков,

таких, что их итерация также является детерминированным языком. Получен критерий и построен автомат с магазинной памятью.

**Теорема 15.** Пусть язык  $L^*$  допускается детерминированным автоматом с магазинной памятью  $\mathfrak{A} = (Q, \{0, 1\}, \Gamma, Z, \Phi, q_0, F)$ . Множество  $L^\infty$  является частично угадываемым тогда и только тогда, когда для некоторой пары  $(q, A), q \in Q, A \in \Gamma$  существует только одна команда вида  $(q, a, A \rightarrow r, \beta), a \in \{0, 1\}, r \in Q, \beta \in \Gamma^*$ .

Рассмотрим класс детерминированных автоматов с магазинной памятью с  $m$  состояниями,  $n$  магазинными символами и максимальной длиной команды  $l$  (под длиной подразумевается число магазинных символов в правой части команды), каждый из которых допускает такой язык  $L^*$ , что  $L^\infty$  является частично угадываемым. Обозначим этот класс  $\widetilde{M}(n, m, l)$ . Пусть  $L(\mathfrak{B})$  — язык, допускаемый автоматом  $\mathfrak{B}$ .

**Теорема 16.** Для любых  $n, m, l \geq 2$  верно

$$\inf_{\mathfrak{B} \in \widetilde{M}(n, m, l)} \sup_{\mathfrak{A} \in \mathfrak{M}} c^{\mathfrak{A}}(L(\mathfrak{B})^\infty) = \frac{1}{l^{n-1}((n-1)(m-1)+1) + m - 1 + \frac{l^{n-1}-1}{l-1}}.$$

## БЛАГОДАРНОСТИ

Автор выражает глубокую благодарность своему научному руководителю профессору Гасанову Э.Э. за постановку задачи и помощь в работе, академику Кудрявцеву В.Б. за многочисленные полезные советы на всех этапах подготовки диссертации и всему коллективу кафедры математической теории интеллектуальных систем механико-математического факультета Московского государственного университета имени М.В.Ломоносова за теплую творческую атмосферу.

## ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации автора по данной теме из списка ВАК:

1. Мастихина А.А. О частичном угадывании сверхслов // Интеллектуальные системы. — 2007. — Т.11, вып.1-4 — С.609-619.
2. Мастихина А.А. Критерий частичного предвосхищения общерегулярных свехсобытий // Дискретная математика. — 2011.—Т.23, вып.4 — С.103-114.

3. Мاستихина А.А. Частичное угадывание сверхсобытий, порожденных простыми  $LL(1)$ -грамматиками // Интеллектуальные системы. — 2011. — Т.15, вып.1-4 — С.507-532.

Публикации не из списка ВАК:

4. Мастихина А.А. Частичное угадывание сверхсобытий, образованных детерминированными контекстно-свободными языками. Материалы X международной конференции "Интеллектуальные системы и компьютерные науки— 2011. — С.157-160.

5. Мастихина А.А. О частичном угадывании сверхслов. Международная конференция "Современные проблемы математики, механики и их приложений", посвященная 70-летию академика В.А.Садовнического (30 марта Ц 2 апреля 2009 г., Москва). Материалы конференции. С. 365-366.

6. Мастихина А.А. Частичное угадывание регулярных выражений. Международная конференция студентов, аспирантов и молодых ученых к Ломоносов-2009нь (2009, Москва). С.45-46.

7. Мастихина А.А. О частичном угадывании регулярных выражений. Международный семинар "Дискретная математика-2010"(2010, Москва). Издательство механико-математического факультета МГУ, Москва, 2010. 385-387.

8. Мастихина А.А. О частичном угадывании множеств сверхслов, заданных общерегулярными выражениями. Международная конференция студентов, аспирантов и молодых ученых "Ломоносов-2010"(2010, Москва).

9. Мастихина А.А. Частичное угадывание некоторых контекстно-свободных языков. Международная конференция студентов, аспирантов и молодых ученых "Ломоносов-2011"(2011, Москва).

10. Мастихина А.А. Частичное предвосхищение множеств последовательностей. Труды 54-й научной конференции МФТИ (10-30 ноября, 2011). Т.1. М.:МФТИ, 2011. С.60-61.