Прогнозирование общерегулярных сверхсобытий автоматами

Э. Э. Гасанов, А. А. Мастихина

В статье доказаны критерий прогнозируемости общерегулярных сверхсобытий и критерий прогнозируемости сверхитераций регулярных событий. Приведены алгоритмы построения прогнозирующих автоматов.

Ключевые слова: прогнозирование событий, автоматы, общерегулярные события.

Введение

В данной работе 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 видеоизображений, фильтры шумоподавления и т.п.

Понятие автомата с предвосхищением появилось еще в начале развития теории автоматов [1]. 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, когда предполагаемый символ не совпал с поступившим на вход на самом деле, происходит откат алгоритма, и выигрыша во времени не происходит.

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 [2].

Также можно использовать для предсказания некоторую информацию о строении последовательности. Например, если она порождена некоторой формальной грамматикой. Наиболее популярны регулярные и контекстно-свободные грамматики, они используются в языках программирования и относительно просты для разбора. Они рассматривались, в частности, в [3, 4, 5].

Несмотря на широкое и повсеместное использование автоматов и автоматного моделирования [6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21] автомат для целей прогнозирования до последнего времени не использовался. Впервые возможность использования автоматов для прогнозирования событий была показана в статье А.Г.Вереникина и Э.Э.Гасанова [22]. В этой статье впервые были введены прогнозирующие автоматы — конечные автоматы, прогнозирующие сверхслово или множество сверхслов. Здесь и далее сверхсловом называется бесконечная последовательность символов некоторого алфавита, например, это последовательность из нулей и единиц. Автомат прогнозирует сверхслово, если через некоторое конечное время после начала подачи сверхслова он начинает угадывать каждый следующий символ, то есть на выходе в момент времени t выдавать элемент входной последовательности под номером $t+1$. Было показано, что прогнозируемо только множество периодических сверхслов с фиксированным периодом. В работе [23] этот результат был обощен на случай многозначных алфавитов.

Поэтому в [24] было введено понятие частичного прогнозирования, которое имеет место в случае, когда автомат угадывает следующий символ не обязательно в каждый момент времени, но достаточно часто.

Вводится степень прогнозирования как предел отношения числа верно предсказанных символов за время t от начала сверхслова к длине его префикса t при t, устремленном к бесконечности.

В зависимости от того, брать ли нижний или верхний предел в случае, когда 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 прогнозирующего есть возможность самостоятельно выбирать, когда подсчитывать эту долю. Кстати, в этом случае несущественно, что последовательность бесконечная, так как можно вычислять долю в момент окончания слова. Здесь мы будем рассматривать частичную прогнозируемость только в смысле нижней степени.

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

Задача состоит в нахождении критериев, позволяющих определить, является ли некоторое сверхсобытие прогнозируемым или частично прогнозируемым. В случае положительного ответа необходимо построить автомат, прогнозирующий или частично прогнозирующий данное сверхсобытие.

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

Результаты могут быть использованы в программировании при создании оптимизирующих компиляторов. Регулярные грамматики широко распространены в этой области, так как они обладают широкими выразительными возможностями, но при этом достаточно просты для распознавания. Прогнозирующие 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 рассматриваемых множеств критерии разграничивают случаи, когда имеет смысл предсказывать, основываясь на поступившей части информационного потока и общей структуре множества, и когда никакой автомат не сможет дать результата, лучшего, чем подбрасывание монеты.

Алгоритмы, предложенные для частичного прогнозирования, 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 прогнозируют существенную долю символов, иногда максимально возможную.

Процесс частичного прогнозирования удобно рассматривать как альтернативную игру двух игроков. Имеется некоторое сверхсобытие L в алфавите A из некоторого класса (в нашем случае из класса общерегулярных сверхсобытий в алфавите $\{0,1\}$), и оба игрока знают сверхсобытие L. Игра идет во времени, и каждый такт каждый игрок делает по одному ходу, причем в каждый момент времени первым ход делает первый игрок, а затем ходит второй игрок. Игра, вообще говоря, идет бесконечно долго. В начальный (первый) момент времени первый игрок, представляющий собой прогнозирующее устройство из некоторого класса (в данном разделе из класса конечных инициальных автоматов), называет некоторый символ $a(1)$ из A, стараясь предугадать символ, который будет назван вторым игроком. Затем ход делает второй игрок и выдает свой символ $b(1)$ из алфавита A так, чтобы хотя бы одно сверхслово из сверхсобытия L начиналось на символ $b(1)$, при этом второй игрок уже знает ход первого игрока. Пусть прошло t тактов времени, и за это время первый игрок сделал t ходов, которые образуют слово $\alpha_t = a(1)\dots a(t)$, и второй игрок сделал t ходов, которые образуют слово $\beta_t = b(1)\dots b(t)$. В $(t+1)$-й момент времени первый игрок выбирает символ $a(t+1)$. После этого второй игрок выбирает символ $b(t+1)$ так, чтобы в сверхсобытии L нашлось сверхслово, начинающееся на слово β_{t+1}. Если оказывается, что $b(t+1) = a(t+1)$, то говорят, что произошло угадывание. Пусть за t тактов произошло d_t угадываний. Задача первого игрока — увеличивать число угадываний d_t, а задача второго игрока — уменьшать это значение. Если в пределе при t стремящемся к бесконечности величина d_t больше нуля, то считается, что первый игрок победил, и прогнозирующее устройство построено, в противном случае побеждает второй игрок, который доказывает, что прогнозирующее устройство для сверхсобытия L построить нельзя.

В статье обобщаются результаты работы [25] и проводятся их тщательные доказательства. В частности, доказан критерий частичной прогнозируемости общерегулярных сверхсобытий в алфавите $\{0,1\}$

конечными инициальными автоматами. Приведены алгоритмы построения прогнозирующих автоматов, продемонстрированные на примерах. Для сверхитераций регулярных событий доказан другой критерий частичной прогнозируемости, который порождает новый алгоритм построения прогнозирующих автоматов.

Основные понятия и формулировка результатов

В данной работе мы исследуем частичное прогнозирование регулярных сверхсобытий в алфавите $\{0,1\}$ конечными инициальными автоматами с входным и выходным алфавитами $\{0,1\}$ (класс таких автоматов обозначим $\mathfrak{M}$).

Пусть $E_2=\{0,1\}$ — конечный алфавит. Через E_2^* и E_2^∞ обозначим соответственно множество всех слов конечной длины и множество всех сверхслов в алфавите E_2.

Если a — сверхслово в алфавите E_2, t — натуральное число, то через $a\underset{t}{]}$ обозначим префикс длины t сверхслова a, т.е. $a\underset{t}{]} = a(1)a(2)\ldots a(t)$.

В соответствии с [26] будем рассматривать конечные автоматы следующего вида:

$$V=(E_2,Q,E_2,\varphi,\psi,q_0)\,,$$

где E_2 — входной алфавит, Q — множество состояний, которое является конечным подмножеством некоторого фиксированного счётного множества, E_2 — выходной алфавит, $\varphi:Q\times E_2\to Q$ — функция переходов, $\psi:Q\times E_2\to E_2$ — функция выходов, q_0 из Q — начальное состояние.

Расширим функции φ и ψ на $Q\times E_2^*$, а именно, если $\alpha\in E_2^*$, $a\in E_2$, то индуктивно определим

$$\varphi(q,\alpha a)=\varphi(\varphi(q,\alpha),a),$$

$$\psi(q,\alpha a)=\psi(\varphi(q,\alpha),a).$$

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

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

$$d^V(\alpha,t) = \frac{1}{t}\sum_{i=1}^{t}(1 - |y^V_\alpha(i) - \alpha(i+1)|).$$

Величина $d^V(\alpha,t)$ равна доле правильно предсказанных выходных символов инициальным автоматом V при подаче на его вход первых t символов сверхслова α.

Инициальный автомат V *прогнозирует сверхслово* α *из* $\{0,1\}^\infty$ *со степенью* $\sigma \in [0,1]$, если

$$c^V(\alpha) = \varliminf_{t\to\infty} d^V(\alpha,t) = \sigma.$$

Величину $c^V(\alpha)$ будем называть *степенью прогнозирования сверхслова* α *инициальным автоматом* V.

Инициальный автомат V *частично прогнозирует множество сверхслов* L, если найдется такое положительное число σ, что для всех α из L выполнено $c^V(\alpha) > \sigma$. Множество L *частично прогнозируемо*, если найдется частично прогнозирующий его инициальный автомат V из $\mathfrak{M}$.

Степень прогнозирования множества L инициальным автоматом V — это

$$c^V(L) = \inf_{\alpha\in L} c^V(\alpha).$$

Множество F, $F \subseteq Q$, автомата $V = (A,Q,B,\varphi,\psi)$ (или автомата без выходов $V = (A,Q,\varphi)$) назовем *сильно связным*, если для любых $q', q'' \in F$ найдется такое слово α из A^*, что $\varphi(q',\alpha) = q''$ и $\overline{\varphi}(q',\alpha) \in F^*$.

Напомним, что если для состояний $q', q'' \in Q$ найдется такое слово α из A^*, что $\varphi(q',\alpha) = q''$, то говорят, что состояние q'' *достижимо* из состояния q'. В дальнейшем будем считать, что для инициальных автоматов все состояния автомата достижимы из начального состояния, поскольку если выбросить недостижимые состояния, то выходная функция автомата не изменится.

Сильно связной компонентой автомата $V = (A,Q,B,\varphi,\psi)$ (или автомата без выходов $V = (A,Q,\varphi)$) назовем максимальное по включению сильно связное множество состояний этого автомата, т.е. такое сильно связное множество состояний, что добавление к нему любого не входящего в него состояния нарушает свойство сильно связности.

Понятно, что у автомата может быть несколько сильно связных компонент.

Назовем сильно связную компоненту *замкнутой*, если из нее не достижима никакая другая сильно связная компонента. Понятно, что замкнутое множество состоит из состояний, из которых все стрелки в диаграмме Мура ведут в состояния из этого же множества. Это свойство замкнутых множеств при наличии свойства сильной связности можно было бы использовать в качестве определения замкнутого множества, откуда сразу же следовало бы, что замкнутое сильно связное множество состояний является сильно связной компонентой.

Очевидно, что в любом конечном автомате хотя бы одна замкнутая сильно связная компонента существует.

Будем рассматривать инициальный автомат без выходов $V = (E_2, Q, \varphi, q_0)$, представляющий некоторое общерегулярное множество $L \subseteq E_2^\infty$ с помощью $F \subset 2^Q$.

Обозначим через $\{\widetilde{Q}_1, ..., \widetilde{Q}_w\}$ множество замкнутых сильно связных компонент автомата V. Понятно, что $w \geqslant 1$. Возможны два случая. Либо в автомате найдется такая замкнутая сильно связная компонента $\widetilde{Q}_i$, что множество ее состояний совпадает с некоторым элементом F, либо такой замкнутой сильно связной компоненты не найдется. Оказывается, что именно это свойство является критериальным при определении частичной прогнозируемости сверхсобытий.

Справедливы следующие теоремы.

Теорема 1. *Если общерегулярное сверхсобытие $L \subseteq E_2^\infty$ представимо в конечном инициальном автомате $V = (E_2, Q, \varphi, q_0)$ с помощью некоторого семейства F, $F \subset 2^Q$, то сверхсобытие L не является частично прогнозируемым тогда и только тогда, когда хотя бы для одной замкнутой сильно связной компоненты $\widetilde{Q}$ автомата V выполнено $\widetilde{Q} \in F$.*

Теорема 2. *Пусть общерегулярное сверхсобытие R^∞, где R — регулярное событие из E_2^*, представимо в конечном инициальном автомате V, $V = (E_2, Q, B, \varphi, \psi, q_0)$ с помощью некоторого семейства F, $F \subset 2^Q$, тогда следующие утверждения эквивалентны:*

1) *сверхсобытие R^∞, не является частично прогнозируемым;*

2) *для любого слова* α *из* $\{0,1\}^*$ *существует слово* β *из* $\{0,1\}^*$ *и сверхслово* γ *из* $\{0,1\}^\infty$, *что* $\beta\alpha\gamma \in R^\infty$;

3) *хотя бы для одной замкнутой сильно связной компоненты* $\widetilde{Q}$ *автомата* $\mathfrak{A}$ *выполнено* $\widetilde{Q} \in F$.

Доказательство теоремы 1

Лемма 1. *Если инициальный автомат* $V = (E_2, Q, \varphi, q_0)$ *представляет некоторое общерегулярное сверхсобытие* $L \subseteq E_2^\infty$ *с помощью семейства* F, $F \subset 2^Q$, *и в автомате* V *найдется такая замкнутая сильно связная компонента* $\widetilde{Q}$, *что* $\widetilde{Q} \in F$, *то сверхсобытие* L *не является частично прогнозируемым.*

Доказательство. Чтобы доказать частичную непрогнозируемость сверхсобытия L, мы для произвольного инициального автомата $G = (\{0,1\}, Q', \{0,1\}, \varphi', \psi', q_0')$ построим сверхслово $\alpha \in L$, для которого $c^G(\alpha) = 0$. Поскольку мы знаем, что множество $\widetilde{Q}$ принадлежит F, мы выберем такое сверхслово α, что $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q}$, т.е. такое сверхслово, при подаче которого с какого-то момента автомат V начинает ходить только по состояниям из $\widetilde{Q}$, причем по всем этим состояниям.

Обозначим $|Q| = n$. Из условия леммы следует, что в автомате V существует замкнутая сильно связная компонента $\widetilde{Q} = \{q_1', q_2', ..., q_m'\} \in F$. В силу сильной связности множества $\widetilde{Q}$ для любого состояния $q_j' \in \widetilde{Q}$ найдется такое слово γ_j длины меньшей $2m < 2n$, что множество пройденных состояний $\overline{\varphi}(q_j', \gamma_j)$ есть все множество $\{q_1', ..., q_m'\}$ и $\varphi(q_j', \gamma_j) = q_j'$.

Замкнутая сильно связная компонента $\widetilde{Q} \in F$ достижима из начального состояния автомата V по некоторому слову $\beta \in \{0,1\}^*$, т.е. $\varphi(q_0, \beta) = q' \in \widetilde{Q}$. Началом нашего сверхслова α будет слово β, которое приводит автомат V в компоненту $\widetilde{Q}$.

Поскольку $\widetilde{Q}$ замкнутая, то для каждого состояния из $\widetilde{Q}$ обе стрелки в диаграмме Мура автомата V, исходящие из данного состояния, ведут в состояния из $\widetilde{Q}$. Поэтому если в момент t, $t > |\beta|$, автомат G выдает символ a, то определяем $\alpha(t+1) = \neg a$, и значит угадывания в момент t не происходит. Таким образом мы можем строить сколь угодно длинные отрезки сверхслова α, где не будет ни одного угадывания, при этом автомат V все время будет оставаться в компоненте

$\widetilde{Q}$. Но при этом нет гарантии, что на этих полностью непрогнозируемых отрезках автомат V обойдет все состояния из $\widetilde{Q}$. Поэтому время от времени, но все реже и реже, мы будем вставлять в сверхслово α слова γ_j, $j \in \{1, \dots m\}$, которые позволяют обойти все состояния из $\widetilde{Q}$. В результате мы с одной стороны добьемся, чтобы доля угаданных символов стремилась у нулю, но чтобы при этом каждое состояние из $\widetilde{Q}$ автомат V проходил бесконечное число раз.

Перейдем к более строгому определению выше описанной процедуры построения сверхслова α.

Если определить входную букву в момент $t+1$ следующим образом $\alpha(t+1) = \neg\psi'(q_0', \underset{t}{\alpha]})$, то в момент t угадывания не произойдет. Таким образом можно строить сколь угодно длинные полностью непрогнозируемые последовательности.

Построим требуемое сверхслово следующим образом: $\alpha = \beta\alpha_1\gamma_{j_1}\alpha_2\gamma_{j_2}...\alpha_i\gamma_{j_i}...$, где α_i — слово длины $i \cdot 2n$, полностью непрогнозируемое для автомата G, находящегося в состоянии $\varphi'(q_0', \beta\alpha_1\gamma_{j_1}...\alpha_{i-1}\gamma_{j_{i-1}})$, а слово γ_{j_i} построено так, чтобы из состояния $q_{j_i} = \varphi(q_0, \beta\alpha_1\gamma_{j_1}...\alpha_i)$ автомат проходил бы по этому слову по всем состояниям из $\widetilde{Q} = \{q_1', ..., q_m'\}$).

Понятно, что построенное слово $\alpha \in L$, поскольку $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q} \in F$.

Подсчитаем степень прогнозирования для данного сверхслова $\widetilde{\alpha}$

$$c^G(\alpha) \leqslant \lim_{t\to\infty} \frac{|\beta| + \sum_{i=1}^t |\gamma_{j_i}|}{\sum_{i=1}^t |\gamma_{j_i}| + \sum_{i=1}^t |\alpha_i| + |\beta|} \leqslant$$
$$\leqslant \lim_{t\to\infty} \frac{2n + t\cdot 2n}{2n(1+2+...+t) + t\cdot 2n + 2n} =$$
$$= \lim_{t\to\infty} \frac{t+1}{\frac{t(t+1)}{2} + t + 1} = \lim_{t\to\infty} \frac{2}{t+2} = 0.$$

Лемма доказана. □

Лемма 2. *Если инициальный автомат без выходов* $V = (E_2, Q, \varphi, q_0)$ *представляет некоторое общерегулярное сверхсобытие* $L \subseteq E_2^\infty$ *с помощью семейства* F, $F \subset 2^Q$, *и в автомате* V *нет такой замкнутой сильно связной компоненты* $\widetilde{Q}$, *что* $\widetilde{Q} \in F$, *то сверхсобытие* L *является частично прогнозируемым.*

Доказательство. Пусть инициальный автомат без выходов $V = (\{0,1\}, Q, \varphi, q_0)$ принимает множество L с помощью семейства подмножеств состояний $F = \{\widetilde{Q}_1, \widetilde{Q}_2, ..., \widetilde{Q}_k\} \subseteq 2^Q$, причем первые l, $l \leqslant k$, множеств максимальны по вложению, и тем самым не входят одно в другое, а для любого множества $\widetilde{Q}_i$, $i \in \{l+1, \dots k\}$, существует такое множество $\widetilde{Q}_j$, $j \in \{1, \dots, l\}$, что $\widetilde{Q}_i \subset \widetilde{Q}_j$. Можно считать, что для любого множества $\widetilde{Q}_i \in F$, $i \in \{1, \dots, k\}$, существует такое сверхслово $\alpha \in L$, что $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q}_i$. В противном случае такое множество $\widetilde{Q}_i$ можно выбросить из F, и оставшееся множество все равно будет представлять сверхсобытие L. Свойство, что $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q}_i$, $i = 1, \dots, k$, означает, что, начиная с некоторого момента времени (этот момент будем называть *моментом стабилизации*), автомат V на сверхслове α будет ходить только по состояниям из $\widetilde{Q}_i$, причем посещать каждое состояние из $\widetilde{Q}_i$ бесконечное число раз. Последнее означает, что $\widetilde{Q}_i$ — сильно связное множество состояний.

Сначала неформально опишем идею алгоритма предсказывания последующих входных символов.

Поскольку каждое множество $\widetilde{Q}_i$, $i \in \{1, \dots, k\}$, является незамкнутым, то существует такое состояние $q'_i \in \widetilde{Q}_i$ и символ $a_i \in \{0,1\}$, что $\varphi(q'_i, a_i) \notin \widetilde{Q}_i$, т.е. одна из стрелок на диаграмме Мура, исходящая из состояния q'_i, выводит из множества $\widetilde{Q}_i$. Предположим, что для некоторого сверхслова α выполнено $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q}_i$, и пусть N — момент стабилизации, т.е. такой момент времени, что для любого $t \geqslant N$ выполняется $\varphi(q_0, \alpha]_t) \in \widetilde{Q}_i$, иными словами после момента N автомат V на сверхслове α ходит только по состояниям из $\widetilde{Q}_i$. Тогда если в некоторый момент t, $t \geqslant N$, выполнено $\varphi(q_0, \alpha]_t) = q'_i$, то в следующий момент автомат обязательно должен пойти по стрелке с отметкой $\neg a_i$, т.е. $\alpha(t+1) = \neg a_i$, и, значит, в этот момент мы однозначно можем предсказать следующий входной символ. Теперь осталось обеспечить, чтобы автомат V достаточно часто заходил в состояние q'_i. Для произвольного состояния $q \in \widetilde{Q}_i$, $q \neq q'_i$, мы можем найти кратчайшее слово $\beta = b(1) \dots b(m)$, что $\varphi(q, \beta) = q'_i$. Тогда если автомат V в некоторый момент t, $t \geqslant N$, находится в состоянии q, то мы можем сказать, что следующим символом будет символ $\neg b(1)$. Если окажется, что $\alpha(t+1) = \neg b(1)$, то произойдет угадывание, в противном случае автомат V перейдет в состояний $q' = \varphi(q, b(1))$, которое находится ближе к состоянию q'_i, и $\varphi(q', b(2) \dots b(m)) = q'_i$. Далее мы предскажем, что следующим символом будет $\neg b(2)$, и опять любо угадаем следующий выходной символ, либо приблизимся к состоянию q'_i. Таким образом, мы своими предсказаниями как бы загоняем автомат V в состояние, где мы гарантируем угадывание. Тем самым либо мы за m тактов попадем в состояние q'_i, где обязательно произойдет угадывание, либо угадывание произойдет на каком-то из предыдущих тактов. Учитывая, что $m \leqslant |\widetilde{Q}_i|$, то в каждые $|\widetilde{Q}_i|$ тактов мы можем обеспечить хотя бы одно угадывание.

Если бы каждая пара множеств из F не пересекалась по состояниям, то мы могли бы реализовать описанную выше процедуру для каждого множества $\widetilde{Q}_i$, $i \in \{1, \ldots, k\}$, независимо, поскольку каждое состояние автомата V входило бы только в одно множество из F, и мы, находясь в произвольном из состояний, всегда бы знали в каком из множеств из F мы находимся, и, значит, знали бы предел сверхслова, которое приводит в данное состояние бесконечное число раз. Тем самым, мы могли бы обеспечить хотя бы одно угадывание каждые $\max_{1 \leqslant i \leqslant k} |\widetilde{Q}_i|$ тактов.

Если для некоторых индексов i и j, $i, j \in \{1, \ldots, k\}$, выполнено $\widetilde{Q}_i \cap \widetilde{Q}_j \neq \emptyset$, то может оказаться, что $q'_i = q'_j$, но при этом $a_i \neq a_j$, т.е. состояния, из которых можно выйти из своих множеств, у множеств $\widetilde{Q}_i$ и $\widetilde{Q}_j$ одинаковые, но стрелки, по которым происходит выход, разные. В этом случае поскольку мы не знаем для произвольного слова $\alpha \in L$ какому из множеств равен предел сверхслова $\overline{\varphi}(q_0, \alpha)$, то мы не сможем сделать правильное предсказание в тот момент, когда автомат V оказывается в состоянии q'_i. Чтобы преодолеть это препятствие, мы изменим автомат V так, чтобы хотя бы максимальные по вложению множества из F не пересекались.

Построим эквивалентный V инициальный автомат $G = (\{0,1\}, Q', \varphi', q'_0)$, принимающий то же множество L с помощью $F' = \{\widetilde{Q}'_1, \widetilde{Q}'_2, ..., \widetilde{Q}'_{k'}\}$, $k' \geqslant k$, состоящее из непересекающихся множеств $\widetilde{Q}'_1, \widetilde{Q}'_2, ..., \widetilde{Q}'_l$, $\widetilde{Q}'_i \cap \widetilde{Q}'_j = \emptyset$, $i \neq j$, $i, j \in \{1, ..., l\}$, и некоторых их подмножеств $\widetilde{Q}'_j \subset \widetilde{Q}'_i$, $j = l+1, ..., k'$, $i \in \{1, 2, ..., l\}$.

Предварительно исключим из рассмотрения те элементы F, которые являются подмножествами других элементов F, т.е. сначала будем рассматривать только первые l элементов F.

Пусть состояние q принадлежит $\widetilde{Q}_{i_1} \cap \widetilde{Q}_{i_2} \cap \ldots \cap \widetilde{Q}_{i_s}$, $i_1 < i_2 < \ldots < i_s \leqslant l$, $s \geqslant 1$. Тогда сопоставим состоянию q состояния $q^{i_1}, q^{i_2}, \ldots, q^{i_s}$, и в каждом множестве $\widetilde{Q}_{i_j}$, $j = 1, 2, ..., s$, заменим q на q^{i_j}. Размножим таким образом каждое состояние, находящееся в максимальных по вложению элементах F. Новые получившиеся множества обозначим $\widetilde{Q}'_i$, $i = 1, 2, ..., l$.

Прежде чем определить переходы между состояниями автомата G, определим на множестве индексов $\{1, 2, ..., l\}$ отношение следования, при этом если $i < l$, то после индекса i следует индекс $i + 1$, а после индекса l следует индекс 1, т.е. в соответствии с данным отношением следования индексы зациклены.

Переходы в автомате G в основном будут соответствовать переходам в автомате V, но при наличии нескольких вариантов (состояние размножено) переход будет осуществляться в состояние из того же множества $\widetilde{Q}'_i$, в котором автомат G находится.

Более точно, пусть в исходном автомате V выполнялось $\varphi(q_1, a) = q_2$, и состояние q_1 было разделено на $q_1^{i_1}, q_1^{i_2}, ..., q_1^{i_m}$, $i_1 < i_2 < \ldots < i_m \leqslant l$, $m \geqslant 1$, а q_2 было разделено на $q_2^{j_1}, q_2^{j_2}, ..., q_2^{j_p}$, $j_1 < j_2 < \ldots < j_p \leqslant l$, $p \geqslant 1$,. Тогда положим $\varphi'(q_1^r, a) = q_2^r$ (по построению $q_1^r, q_2^r \in \widetilde{Q}'_r$) для всех $r \in \{i_1, i_2, ..., i_m\} \cap \{j_1, j_2, ..., j_p\}$, и $\varphi'(q_1^r, a) = q_2^{j_s}$, если $r \in \{i_1, i_2, ..., i_m\} \backslash \{j_1, j_2, ..., j_p\}$, где j_s — ближайший индекс к индексу r такой, что $j_s > r$, при этом если $r > j_p$, то $j_s = j_1$ (напомним, что в нашем отношении следования после последнего индекса следует первый).

Множество F' образуется из F заменой всех максимальных по вложению множеств $\widetilde{Q}_i$ на соответствующие $\widetilde{Q}'_i$, $i = 1, 2, \ldots, l$. Иными словами множество состояний $\widetilde{Q}'_i = \{q_1^i, q_2^i, \ldots, q_m^i\}$ является элементом F' тогда и только тогда, когда соответствующее множество $\widetilde{Q}_i = \{q_1, q_2, \ldots, q_m\} \in F$, $i = 1, 2, \ldots, l$. Теперь разберемся с элементами F, являющимися подмножествами других элементов F, т.е. с элементами $\widetilde{Q}_i$, где $i \in \{l+1, \ldots, k\}$. Если некоторое множество $\widetilde{Q}_i = \{q_1, \ldots, q_m\} \in F$, $i \in \{l+1, \ldots, k\}$, и $\{q_1, \ldots, q_m\} \subset \widetilde{Q}_{j_1} \cap \widetilde{Q}_{j_2} \cap \ldots \cap \widetilde{Q}_{j_p}$, $j_1 < j_2 < \ldots < j_p \leqslant l$, $p \geqslant 1$, то в множество F' включается p множеств $\{q_1^{j_r}, \ldots, q_m^{j_r}\}$, $r = 1, \ldots, p$. Поэтому количество элементов F' может быть больше, чем количество элементов F. По построению в множестве F' все максимальные по включению элементы не пересекаются.

Положим $Q' = (Q \setminus (\cup_{R \in F} R)) \cup (\cup_{R \in F'} R)$.

Если состояние q_0 из Q было разделено на $q_0^{i_1}, q_0^{i_2}, \ldots, q_0^{i_m}$, $i_1 < i_2 < \ldots < i_m \leqslant l$, $m \geqslant 1$, то в качестве q_0' примем $q_0^{i_1}$, в противном случае положим $q_0' = q_0$.

Пусть L' — сверхсобытие, которое представляет полученный автомат $G = (\{0,1\}, Q', \varphi', q_0')$ с помощью множества F'.

Покажем, что $L' = L$.

Пусть $\alpha \in L$, т.е. для некоторого i, $i \in \{1, \ldots, k\}$, выполняется $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q}_i \in F$, и пусть, начиная с момента стабилизации N, автомат V не выходит за множество состояний $\widetilde{Q}_i$, т.е. для любого $t \geqslant N$ выполняется $\varphi(q_0, \alpha]_t) \in \widetilde{Q}_i$. Пусть $\varphi(q_0, \alpha]_N) = q$, и пусть $q \in \widetilde{Q}_{j_1} \cap \widetilde{Q}_{j_2} \cap \ldots \cap \widetilde{Q}_{j_m}$, $j_1 < j_2 < \ldots < j_m \leqslant l$, $m \geqslant 1$. Пусть $\varphi'(q_0', \alpha]_N) = q'$. По построению функции φ' понятно, что $q' \in \{q^{j_1}, q^{j_2}, \ldots, q^{j_m}\}$.

Предположим сначала, что $\widetilde{Q}_i$ не является подмножеством никакого другого множества из F, т.е. $i \leqslant l$. Поскольку $q \in \widetilde{Q}_i$, то $i \in \{j_1, j_2, \ldots, j_m\}$.

Если $q' = q^i$, т.е. $q' \in \widetilde{Q}_i'$, то поскольку автомат V все время после момента N остается в множестве $\widetilde{Q}_i$ и проходит каждое из состояний из $\widetilde{Q}_i$ бесконечное число раз, то по построению функции φ' автомат G будет все время оставаться в множестве $\widetilde{Q}_i'$, причем будет заходить в каждое из состояний из $\widetilde{Q}_i'$ бесконечное число раз. Значит, $\lim(\overline{\varphi'}(q_0', \alpha)) = \widetilde{Q}_i' \in F'$.

Рассмотрим теперь случай, когда $q' \neq q^i$. Будем считать, что $q' = q^{j_s}$ и $j_s < i$.

Так как $\widetilde{Q}_i$ и $\widetilde{Q}_{j_s}$ — максимальные по вложению множества, то $\widetilde{Q}_i \setminus \widetilde{Q}_{j_s} \neq \emptyset$. Поскольку автомат V проходит все состояния из $\widetilde{Q}_i$, то наступит такой момент t_1, $t_1 > N$, когда автомат V в первый раз после момента N выйдет из множества $\widetilde{Q}_{j_s}$ и перейдет в такое состояние q_1, что $q_1 \in \widetilde{Q}_i \setminus \widetilde{Q}_{j_s}$, причем в момент $t_1 - 1$ автомат V находился в состоянии q_1' и $q_1' \in \widetilde{Q}_{j_s}$, т.е. $\varphi(q_1', \alpha(t_1 - 1)) = q_1$. Пусть $q_1 \in \widetilde{Q}_{r_1} \cap \widetilde{Q}_{r_2} \cap \ldots \cap \widetilde{Q}_{r_m}$, $r_1 < r_2 < \ldots < r_m$. Понятно, что $i \in \{r_1, r_2, \ldots, r_m\}$, но $j_s \notin \{r_1, r_2, \ldots, r_m\}$. Пусть r_p ближайший сверху индекс к индексу j_s. Очевидно $r_p \leqslant i$. По построению функции φ' имеем $\varphi'(q^{j_s}, \alpha(t_1 - 1)) = q_1^{r_p}$, т.е. автомат G в момент t_1 переходит в состояние из множества $\widetilde{Q}_{r_p}'$, причем индекс r_p ближе к индексу i, чем индекс j_s. Если $r_p = i$, то автомат G уже никогда не выйдет из множества $\widetilde{Q}_i'$ и $\lim(\overline{\varphi'}(q_0', \alpha)) = \widetilde{Q}_i' \in F'$. Если же $r_p < i$, то поскольку $\widetilde{Q}_i \setminus \widetilde{Q}_{r_p} \neq \emptyset$, то автомат G по аналогии с предыдущим случаем выйдет из множества $\widetilde{Q}'_{r_p}$ и перейдет либо в множество $\widetilde{Q}'_i$, либо в некоторое множество, которое по индексу ближе к i чем r_p. Таким образом автомат G через какое-то время обязательно попадет в множество $\widetilde{Q}'_i$ и уже из него не выйдет, тем самым в случае, когда $\widetilde{Q}_i$ не является подмножеством никакого другого множества из F, имеем $\lim(\overline{\varphi'}(q'_0, \alpha)) = \widetilde{Q}'_i \in F'$. Случай, когда $i < j_s$, доказывается аналогично, но в соответствии с отношением следования индексов сначала индексы множеств будут увеличиваться (приближаясь к i в соответствии с отношением следования, но при этом как бы удаляясь), но, перевалив через максимальный индекс, индексы станут приближаться к i снизу.

Предположим теперь, что $i \in \{l+1, \ldots, k\}$ и $\widetilde{Q}_i = \{q_1, \ldots, q_n\} \subset \widetilde{Q}_{r_1} \cap \widetilde{Q}_{r_2} \cap \ldots \cap \widetilde{Q}_{r_p}$, $r_1 < r_2 < \ldots < r_p \leqslant l$. По тому как мы строили множество F' элементами этого множества являются множества $\{q_1^{r_u}, \ldots, q_n^{r_u}\}$, $u = 1, \ldots, p$. Пусть $q' = q^{j_s}$.

Сначала рассмотрим случай, когда $j_s \in \{r_1, r_2, \ldots, r_p\}$. Поскольку автомат V все время после момента стабилизации N будет ходить по состояниям из множества $\{q_1, \ldots, q_n\}$ и будет заходить в каждое из них бесконечное число раз, то по построению функции φ' автомат G будет бесконечное число раз заходить в состояния из множества $\{q_1^{j_s}, \ldots, q_n^{j_s}\}$, все время оставаясь в этом множестве. Значит, $\lim(\overline{\varphi'}(q'_0, \alpha)) = \{q_1^{j_s}, \ldots, q_n^{j_s}\} \in F'$.

Осталось рассмотреть случай, когда $j_s \notin \{r_1, r_2, \ldots, r_p\}$. Предположим $j_s < r_u$, причем если $u \in \{2, \ldots, p\}$, то $r_{u-1} < j_s$. Поскольку $\widetilde{Q}_i \not\subset \widetilde{Q}_{j_s}$, то $\widetilde{Q}_i \setminus \widetilde{Q}_{j_s} \neq \emptyset$. Следовательно, когда автомат V, обходя множество $\widetilde{Q}_i$, выйдет из множества $\widetilde{Q}_{j_s}$, тогда автомат G выйдет из множества $\widetilde{Q}'_{j_s}$ и перейдет либо в множество $\widetilde{Q}'_{r_u}$, либо в такое множество $\widetilde{Q}'_w$, что $j_s < w < r_u$, причем $\widetilde{Q}_i \setminus \widetilde{Q}_w \neq \emptyset$. Следовательно, продолжая обход автоматом V множества $\widetilde{Q}_i$, мы в конце концов автоматом G попадем в состояние из множества $\widetilde{Q}'_{r_u}$, и так в нем и останемся, т.е. $\lim(\overline{\varphi'}(q'_0, \alpha)) = \{q_1^{r_u}, \ldots, q_n^{r_u}\} \in F'$. Случай, когда $j_s > r_p$, рассматривается аналогично, но в этом случае $\lim(\overline{\varphi'}(q'_0, \alpha)) = \{q_1^{r_1}, \ldots, q_n^{r_1}\} \in F'$.

Тем самым мы доказали, что $L \subseteq L'$.

Возьмем произвольное сверхслово α из L'. Значит, для некоторого i, $i \in \{1, \ldots, k'\}$, выполняется $\lim(\overline{\varphi}(q_0, \alpha)) = \widetilde{Q}'_i = \{q_1^r, q_2^r, \ldots, q_m^r\} \in F'$.

По построению функции φ' имеем, что если $\varphi'(q_j^s, a) = q_l^p$, то $\varphi(q_j) = q_l$. Отсюда сразу следует, что если $\lim(\overline{\varphi}'(q_0', \alpha)) = \{q_1^r, q_2^r, \ldots, q_m^r\}$, то $\lim(\overline{\varphi}(q_0, \alpha)) = \{q_1, q_2, \ldots, q_m\}$. Но мы так строили множества из F', что если $\{q_1^r, q_2^r, \ldots, q_m^r\} \in F'$, то $\{q_1, q_2, \ldots, q_m\} \in F$. Следовательно, $\lim(\overline{\varphi}(q_0, \alpha)) \in F$, т.е. $\alpha \in L$.

Значит, $L = L'$, и мы доказали, что автомат G представляет сверхсобытие L с помощью множества F'.

Теперь построим автомат $G' = (\{0,1\}, Q', \varphi', \psi', q_0')$, частично угадывающий L, добавив функцию выходов ψ' к автомату G.

Множество всех состояний, входящих в какой-нибудь элемент F', обозначим Q'', т.е. $Q'' = \cup_{R \in F'} R$.

Выходную функцию определим как функцию от состояния, в которое приходит автомат G', то есть $\psi'(q, a) = f(\varphi'(q, a))$. Опишем алгоритм определения функции $f : Q' \to \{0,1\}$.

В состояниях не из Q'' функцию f определим произвольным образом. А для состояний из Q'' функция f будет устроена так, чтобы автомат либо угадывал, либо приближался к состояниям, в которых угадывание гарантировано.

1-*й шаг.* Для состояний, не принадлежащих Q'', расставим выходы произвольным образом. Например, для каждого состояния q, $q \notin Q''$, определим $f(q) = 0$. Для каждого максимального по включению множества $\widetilde{Q}_i'$, $i = 1, 2, \ldots, l$, проделаем следующую процедуру. Состояния $q \in \widetilde{Q}_i'$, для которых найдется буква $a \in \{0,1\}$, такая что $\varphi(q, a) \notin \widetilde{Q}_i'$, занесем в класс $Q(1)$. Поскольку $\widetilde{Q}_i'$ не замкнутая компонента, то хотя бы одно такое состояние в $\widetilde{Q}_i'$ найдется. Теперь для каждого состояния $q \in \widetilde{Q}_i' \cap Q(1)$ если $\varphi(q, a) \notin \widetilde{Q}_i'$, то назначим $f(q) = \neg a$. $Q(1)$ — это множество состояний, в которых после момента стабилизации угадывание гарантировано.

i-*й шаг.* Пусть выделены классы $Q(1), ..., Q(i-1)$, и для этих состояний функция f определена. Рассмотрим множество $Q(i) = \{q \in Q'' \setminus \{Q(1) \cup \ldots \cup Q(i-1)\}$: существует такой символ $a \in E_2$, что $\varphi(q, a) \in Q(i-1)\}$. Зададим функцию f на состояниях из $Q(i)$ следующим образом: если для $q \in Q(i)$ и $a \in E_2$ выполнено $\varphi(q, a) \in Q(i-1)$, то $f(q) = \neg a$. При этом, если $\varphi(q, 0) \in Q(i-1)$ и $\varphi(q, 1) \in Q(i-1)$, то значение $f(q)$ можно выбрать произвольно, например, назначить $f(q) = 0$. $Q(i)$ — это множество состояний, из которых за $i-1$ тактов можно попасть в какое-то состояние гарантированного угадывания из $Q(1)$.

Очевидно, что если $\max_{1\leqslant i\leqslant l}|\widetilde{Q}'_i| = n$, то не позднее, чем через n шагов функция f будет полностью определена, т.е. $Q'' = \cup_{i=1}^{n'} Q(i)$, $n' \leqslant n$.

Покажем, что для любого сверхслова α из множества L данный автомат G' после момента стабилизации не может не угадывать n символов подряд.

Отметим, что в состояниях из $Q(i)$ автомат G' предсказывает символ, который в диаграмме Мура соответствует ссылке, ведущей в состояние из $Q(j)$, где $j \geqslant i$, и, значит, если в этом состоянии угадывания не происходит, то автомат G' переходит в состояние из $Q(i-1)$ (из некоторых состояний из $Q(i)$ обе стрелки могут вести в состояния из $Q(i-1)$).

Если $\lim \overline{\varphi}(\alpha) = \widetilde{Q}'_i$, где $i \in \{1, 2, \ldots, l\}$, то после момента стабилизации в состояниях $Q(1)$ автомат всегда угадывает. В остальных состояниях из Q'' по построению автомат G' либо угадывает символ, либо переходит из $Q(i)$ в $Q(i-1)$.

Если $\lim \overline{\varphi}(\alpha) = \widetilde{Q}'_i$, где $i \in \{l+1, \ldots, k'\}$, то для некоторого j, $j \in \{1, 2, \ldots, l\}$, выполняется $\widetilde{Q}'_i \subseteq \widetilde{Q}'_j$. В этом случае может так получится, что $Q(1) \cap \widetilde{Q}'_i = \emptyset$, и автомат G' не попадает в состояния гарантированного угадывания. Но если в $\widetilde{Q}'_i$ есть состояния из $Q(r)$ и нет ни одного состояния из $Q(1), ..., Q(r-1)$, то в состояниях из $Q(r) \cap \widetilde{Q}'_i$ автомат G' будет гарантировано угадывать, поскольку в этих состояниях автомат G' не будет идти по стрелке, которая ведет в состояние из $Q(r-1)$, а пойдет по стрелке, которая ведет в состояние из $\widetilde{Q}'_i$, а именно символ, который приписан этой стрелке автомат G' и предсказывает.

Остается заметить, что из любого состояния путь в диаграмме Мура, ведущий в состояние гарантированного угадывания, имеет длину не большую n. Следовательно, после момента стабилизации каждые n тактов хотя бы одно угадывание будет происходить, и, значит, степень прогнозирования сверхсобытия L автоматом G' не меньше чем $1/n$.

Заметим, что поскольку каждое из множеств $\widetilde{Q}'_i$, $i \in \{1, 2, \ldots, k'\}$, не замкнутое, и следовательно $n < |Q'|$, то степень прогнозирования сверхсобытия L автоматом G' не меньше чем $\frac{1}{|Q'|-1}$. $\square$

Пример 1. Рассмотрим инициальный автомат $V = (E_2, \{q_1, q_2, q_3\}, E_2, \varphi, q_1)$, где функция φ задана таблицей 1. Ав-

Таблица 1

φ	q_1	q_2	q_3
0	q_3	q_1	q_3
1	q_2	q_1	q_3

томат принимает сверхсобытие $L = (11 \cup 10)^\infty$ с помощью множества $F = \{\{q_1, q_2\}\}$. Построить автомат, частично прогнозирующий это сверхсобытие.

Решение. Диаграмма Мура автомата, заданного таблицей 1, приведена на рисунке 1 а). Сильно связные компоненты автомата — $\widetilde{Q}_1 = \{q_1, q_2\}, \widetilde{Q}_2 = \{q_3\}$. Компонента $\widetilde{Q}_2$ является единственной замкнутой и не является элементом F. Назначим в состоянии из $\widetilde{Q}_2$ выход произвольно: $f(q_3) = 0$.

Далее, существует единственная стрелка, выводящая из компоненты $\widetilde{Q}_2$, — это стрелка с входным символом 0, ведущая из состояния q_1, т.е. $\varphi(q_1, 0) = q_3$, поэтому $Q(1) = \{q_1\}$ и $f(q_1) = 1$.

$\varphi(q_2, 0) = \varphi(q_2, 1) = q_1$, положим $Q(2) = \{q_2\}$, выход в этом состоянии может быть любым, пусть $f(q_2) = 1$.

Если перейти на привычный язык, то получаем, что функция выхода ψ задается следующей таблицей 2.

Таблица 2

ψ	q_1	q_2	q_3
0	0	1	0
1	1	1	0

Традиционная диаграмма Мура описанного прогнозирующего автомата приведена на рисунке 1 б). В ней функция выхода $\psi(q, x)$ записывается внутри круга, соответствующего состоянию q. Но более удобной для описания прогнозирующих автоматов представляется диаграмма, когда внутри круга, соответствующего состоянию q, будет записано значение функции $f(q)$. Такие диаграммы, внешне не отличающиеся от обычных диаграмм Мура, но по-другому интерпретируемые, договоримся называть диаграммами Мура второго рода. В

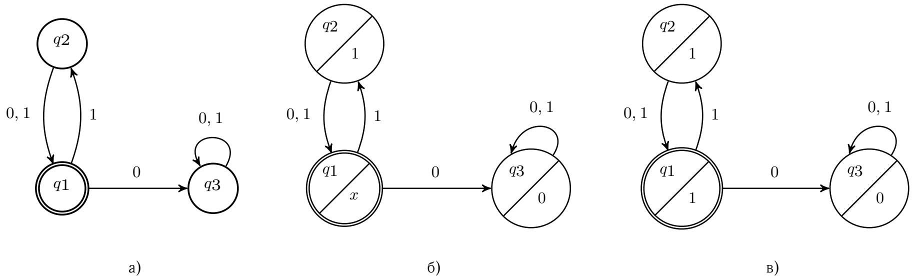

Рис. 1: а) Автомат, в котором сверхсобытие $(11 \cup 10)^\infty$ представимо с помощью $F = \{\{q_1, q_2\}\}$. б) Традиционная диаграмма Мура автомата, частично прогнозирующего это событие. в) Диаграмма Мура второго рода автомата, частично прогнозирующего это событие.

диаграмме Мура второго рода значение функции $\psi(q, a)$ можно найти внутри круга, в который ведет стрелка с пометкой a, ведущая из круга, соответствующего состоянию q. Диаграмма Мура второго рода описанного выше прогнозирующего автомата приведена на рисунке 1 в).

Легко видеть, что на сверхсловах из сверхсобытия $(11 \cup 10)^\infty$ автомат ходит только по состояниям q_1 и q_2, причем когда он находится в состоянии q_1, на вход к нему может поступить только символ 1. Поэтому, находясь в состоянии q_2, из которого мы всегда переходим в состояние q_1, мы смело можем выдавать символ 1, предсказывая, что в следующий момент на вход поступит символ 1. Видно, что сверхслово $(1)^\infty$ хуже всего прогнозируется этим автоматом, и его степень прогнозирования $1/2$.

Из лемм 1 и 2 следует теорема 1.

Пример 2. Автомат, диаграмма Мура, которого приведена на рисунке 2, принимает некоторое сверхсобытие с помощью множества $F = \{\{q_1, q_2, q_3\}, \{q_2, q_3, q_4\}, \{q_2, q_3\}\}$. Построить автомат, частично прогнозирующий это сверхсобытие.

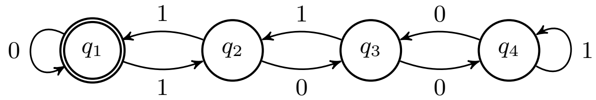

Рис. 2: Автомат, принимающий сверхсобытие с помощью множества $F = \{\{q_1, q_2, q_3\}, \{q_2, q_3, q_4\}, \{q_2, q_3\}\}$

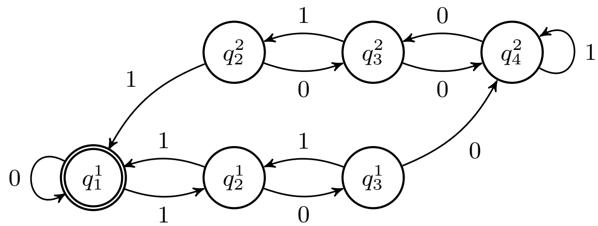

Рис. 3: Автомат, эквивалентный автомату, приведенному на рисунке 2.

Решение. Единственной замкнутой сильно связной компонентой автомата является множество всех состояний $\{q_1, q_2, q_3, q_4\}$, и оно не входит в множество F. Следовательно сверхсобытие, которое представляет данный автомат, является частично прогнозируемым.

Обозначим $\widetilde{Q}_1 = \{q_1, q_2, q_3\}$, $\widetilde{Q}_2 = \{q_2, q_3, q_4\}$, $\widetilde{Q}_3 = \{q_2, q_3\}$. Исключим пока $\widetilde{Q}_3$ из рассмотрения, так как оно содержится в $\widetilde{Q}_1$ и в $\widetilde{Q}_2$.

Состояние q_1 переходит в состояние q_1^1. Состояние $q_2 \in \widetilde{Q}_1 \cup \widetilde{Q}_2$, поэтому ему соответствуют состояния q_2^1, q_2^2. Состояние $q_3 \in \widetilde{Q}_1 \cup \widetilde{Q}_2$, заменяем его на q_3^1, q_3^2. Состояние q_4 заменяем на q_4^2.

В результате получаем непересекающиеся множества $\widetilde{Q}'_1 = \{q_1^1, q_2^1, q_3^1\}$, $\widetilde{Q}'_2 = \{q_2^2, q_3^2, q_4^2\}$. Поскольку $\widetilde{Q}_3 \subseteq \widetilde{Q}_1 \cap \widetilde{Q}_2$, то из $\widetilde{Q}_3$ получаем два множества $\widetilde{Q}'_3 = \{q_2^1, q_3^1\}$ и $\widetilde{Q}''_3 = \{q_2^2, q_3^2\}$.

Переходы автомата $G = (\{0,1\}, \{q_1^1, q_2^1, q_3^1, q_2^2, q_3^2, q_4^2\}, \varphi, q_1^1)$ изображены на рисунке 3. Он эквивалентен автомату, приведенному на рисунке 2, и принимает то же сверхсобытие с помощью множества $F' = \{\widetilde{Q}'_1, \widetilde{Q}'_2, \widetilde{Q}'_3, \widetilde{Q}''_3\}$.

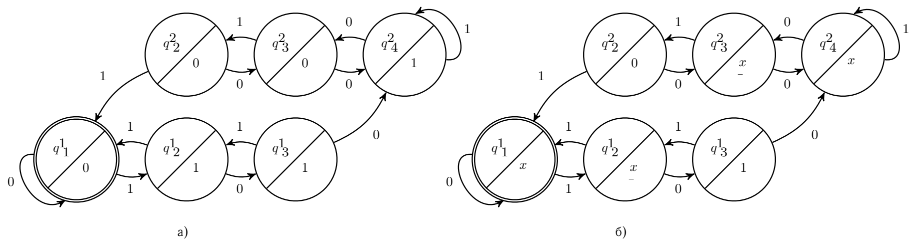

Рис. 4: а) Диаграмма Мура второго рода прогнозирующего автомата. б) Традиционная диаграмма Мура прогнозирующего автомата.

Выйти из множества $\widetilde{Q}'_1$ можно из состояния q_3^1 по стрелке с пометкой 0, следовательно $f(q_3^1) = 1$. Выйти из множества $\widetilde{Q}'_2$ можно из состояния q_2^2 по стрелке с пометкой 1, следовательно $f(q_2^2) = 0$. Значит, $Q(1) = \{q_3^1, q_2^2\}$.

Переходы в состояния из $Q(1)$ осуществляются по стрелкам с пометкой 0 из состояния q_2^1 и с пометкой 1 из q_3^2, следовательно, $Q(2) = \{q_2^1, q_3^2\}$ и $f(q_2^1) = 1$, $f(q_3^2) = 0$.

И, наконец, перейти в состояния из $Q(2)$ можно по стрелкам с пометкой 1 из состояния q_1^1 и с пометкой 0 из q_4^2, следовательно, $Q(3) = \{q_1^1, q_4^2\}$ и $f(q_1^1) = 0$, $f(q_4^2) = 1$.

Диаграмма Мура второго рода описанного прогнозирующего автомата приведена на рисунке 4 а), а традиционная диаграмма Мура этого же автомата на рисунке 4 б).

Доказательство теоремы 2

В случае, когда угадываемое сверхсобытие является сверхитерацией регулярного события, то существует другой критерий частичной прогнозируемости.

Лемма 3. *Если инициальный автомат $V = (E_2, Q, \varphi, q_0)$ представляет общерегулярное сверхсобытие R^∞, где R — регулярное событие из E_2^*, с помощью семейства F, $F \subset 2^Q$, и в автомате*

V найдется такая замкнутая сильно связная компонента $\widetilde{Q}$, что $\widetilde{Q} \in F$, то для любого слова α из E_2^* существует слово β из E_2^* и сверхслово γ из E_2^∞, что $\beta\alpha\gamma \in R^\infty$.

Доказательство. Для произвольного α найдем такие $\beta \in \{0,1\}^*, \gamma \in \{0,1\}^\infty$, что $\beta\alpha\gamma \in R^\infty$. Рассмотрим замкнутую сильно связную компоненту $\widetilde{Q}$. Она достижима из начального состояния по некоторому слову $\beta \in \{0,1\}^*$, т.е. $\varphi(q_0, \beta) = q' \in \widetilde{Q}$. Так как множество $\widetilde{Q}$ замкнутое, то выполнено $\varphi(q', \alpha) = q'' \in \widetilde{Q}$. Так как $\widetilde{Q}$ — сильно связное множество, то существует такое $\delta \in \{0,1\}^*$, что $\varphi(q'', \delta) = q''$ и в слове $\overline{\varphi}(q'', \delta)$ встречаются все состояния из $\widetilde{Q}$. Возьмем $\gamma = \delta^\infty$. Тогда $\lim(\overline{\varphi}(\beta\alpha\gamma)) = \widetilde{Q} \in F$ и, следовательно, $\beta\alpha\gamma \in R^\infty$. Лемма доказана. $\square$

Лемма 4. *Если инициальный автомат $V = (E_2, Q, \varphi, q_0)$ представляет общерегулярное сверхсобытие R^∞, где R — регулярное событие из E_2^*, с помощью семейства F, $F \subset 2^Q$, и в автомате V не найдется такой замкнутой сильно связной компоненты $\widetilde{Q}$, что $\widetilde{Q} \in F$, то найдется слово $\alpha \in \{0,1\}^*$, которое не является подсловом никакого сверхслова из R^∞.*

Доказательство. Пусть $F = \{\widetilde{Q}_1, \widetilde{Q}_2, \ldots, \widetilde{Q}_m\}$. Можно считать, что все состояния из множеств $\widetilde{Q}_i$, $i = 1, 2, \ldots, m$, достижимы из начального состояния q_0. Поскольку если в некотором множестве $\widetilde{Q}_i$ есть недостижимое из q_0 состояние, то в сверхсобытии R^∞ нет ни одного сверхслова, предел которого будет равен $\widetilde{Q}_i$. Следовательно, в этом случае множество $\widetilde{Q}_i$ можно выбросить из F и представлять R^∞ с помощью $F \setminus \{\widetilde{Q}_i\}$.

Сначала покажем, что если для некоторого состояния q, $q \in Q$, выполняется $\varphi(q_0, \beta) = q$ и для любого слова $\gamma \in \{0,1\}^*$ справедливо $\varphi(q, \gamma) \in \widetilde{Q}_i$ для некоторого i из $\{1, 2, \ldots, m\}$, то никакое сверхслово из R^∞ не начинается на слово β. В самом деле, поскольку никакое слово не выводит из множества $\widetilde{Q}_i$, то либо $\widetilde{Q}_i$ — замкнутое множество, чего не может быть по условию леммы, либо некоторое собственное подмножество Q' множества $\widetilde{Q}_i$ (не принадлежащее F) является замкнутым. Отсюда сразу следует, что предел любого сверхслова, начинающегося на слово β, не может принадлежать семейству F, и, значит, это сверхслово не принадлежит R^∞. В частности, если $\beta = \Lambda$, то $R^\infty = \emptyset$, и утверждение леммы справедливо.

Построим такое слово α^1, что $\varphi(q_0, \alpha^1) \notin \widetilde{Q}_1$. (Эти состояния должны быть достижимы, иначе нарушается условие леммы.) Далее для каждого $i = 2, \ldots, m$ найдем такое α^i, что $\varphi(q_0, \alpha^1 \ldots \alpha^i) \notin \widetilde{Q}_i$. Если такого слова не находится, согласно доказанному выше утверждению ни одно сверхслово из R^∞ не начинается на $\alpha^1 ... \alpha^{i-1}$, и в этом случае лемма доказана.

Слово $\alpha_0 = \alpha^1 ... \alpha^m$ либо является искомым, либо найдутся такие слово γ_1 и сверхслово β, что $\gamma_1 \alpha_0 \beta \in R^\infty$. В этом случае найдется и такое слово β_1, что $\gamma_1 \alpha_0 \beta_1 \in R^*$. Пусть $\varphi(q_0, \gamma_1 \alpha_0 \beta_1) = q_1$.

Для этого состояния аналогично построим слово α_1, последовательно приводящее его в достижимые состояния не из множеств $\widetilde{Q}_1, \ldots, \widetilde{Q}_m$. Для него также либо найдутся слова γ_2 и β_2, такие что $\gamma_2 \gamma_1 \alpha_0 \beta_1 \alpha_1 \beta_2 \in R^*$, либо это слово — искомое.

Повторим процесс до тех пор пока не окажемся в уже пройденном состоянии, то есть $\varphi(q_i, \alpha_i \beta_{i+1} \ldots \alpha_j \beta_{j+1}) = q_i$. Тогда сверхслово $\delta = \gamma_{j+1} \ldots \gamma_1 \alpha_0 \beta_1 \ldots \alpha_{i-1} \beta_i (\alpha_i \beta_{i+1} ... \alpha_j \beta_{j+1})^\infty$ по построению принадлежит R^∞. Тогда оно должно представляться данным автоматом с помощью F, т.е. $\lim(\overline{\varphi}(q_0, \delta)) \in F$. Но слово α_j построено так, что для любого $l \in \{1, 2, \ldots, m\}$ в слове $\theta = \overline{\varphi}(q_i, \alpha_i \beta_{i+1} \ldots \alpha_j \beta_{j+1})$ встречается состояние, не принадлежащее $\widetilde{Q}_l$. А поскольку слово θ встречается в сверхслове $\overline{\varphi}(q_0, \delta)$ бесконечное число раз, то $\lim(\overline{\varphi}(q_0, \delta)) \notin F$.

Полученное противоречие означает, что одно из слов $\alpha_0, \gamma_1 \alpha_0 \beta_1 \alpha_1, \ldots, \gamma_j \ldots \gamma_1 \alpha_0 \beta_1 \alpha_1 \ldots \beta_j \alpha_j$ не является подсловом никакого сверхслова из R^∞.

Лемма доказана. □

Как следствие лемм 1-4 получаем теорему 2.

Из утверждения теоремы 2 вытекает другой способ частичного прогнозирования. Пусть известно, что R^∞ не содержит некоторого подслова $a_1 \ldots a_l$. Тогда можно построить автомат, который своим состоянием запоминает слово длины $(l-1)$, состоящее из последних поступивших входных букв. Поэтому если автомат окажется в состоянии $a_1 \ldots a_{l-1}$, то он знает, что в следующий момент поступит символ $\neg a_l$ и угадает. При этом из состояния $a_1 \ldots a_{l-1}$ по стрелке с входным символом a_l автомат переходит в отдельно введенное тупиковое состояние q_T, в которое он никогда не должен попадать на сверхсловах из R^∞. В остальных состояниях выходы автомата формируются так, чтобы в случае неугадывания автомат приближался бы к состоянию

$a_1 \dots a_{l-1}$, где угадывание гарантировано, т.е. мы можем воспользоваться алгоритмом определения выходной функции, описанным в доказательстве леммы 2, но в данном случае выходную функцию можно определить явным образом. Определим данный автомат формально.

Прогнозирующий автомат для сверхсобытия R^{∞}, не содержащего подслова $a_1 \dots a_l$, имеет следующий вид: $V(a_1 \dots a_l) = (E_2, Q, E_2, \varphi, \psi, q_0)$, где

$$Q = \{q_T\} \cup \{q(1)...q(l-1)|q(i) \in E_2, i = 1..l-1\},$$

функция переходов задается соотношениями:

$$\begin{cases} \varphi(q(1)\dots q(l-1), a) = q(2)\dots q(l-1)a, \text{если } q(1)\dots q(l-1)a \neq a1 \dots al{-}1al, \\ \varphi(a1 \dots al{-}1, al) = qT, \\ \varphi(qT, a) = qT, a \in E2; \end{cases}$$

функция выходов имеет вид $\psi(q, a) = f(\varphi(q, a))$, $q \in Q$, $a \in E_2$, и функция f определяется следующим образом: $f(q_T) = 0$; для каждого $q(1) \dots q(l-1)$ найдем наибольшее k, что $q(1) \dots q(l-1) = q(1) \dots q(l-k-1)a_1 \dots a_k$, и определим $f(q(1) \dots q(l-1)) = \neg a_{k+1}$. В частности, если $k = 0$, то $f(q(1) \dots q(l-1)) = \neg a_1$, если $k = l-1$, то $f(a_1 \dots a_{l-1}) = \neg a_l$. Чтобы не свалиться раньше времени в тупиковое состояние, в качестве начального состояния можно взять любое, для которого $k = 0$, например, можно положить $q_0 = \neg a_1 \neg a_1 \dots \neg a_1$.

В автомате такого вида состояние $a_1 ... a_{l-1}$ достижимо из любого другого, кроме тупикового, по пути длины не большей $l-1$. Поэтому не может быть более l неугаданных символов подряд, следовательно, степень прогнозирования автомата $V(a_1 \dots a_l)$ не меньше чем $1/(l-1)$.

Пример 3. Построить автомат, частично прогнозирующий множество всех последовательностей, не содержащих подслово $a_1a_2a_3 = 010$.

Решение. Автомат $V(010)$ будет хранить в памяти два предыдущих входных символа, поэтому множество состояний имеет вид $Q = \{00, 11, 10, 01, q_T\}$.

Определим функцию переходов: $\varphi(q_1q_2, a) = q_2a$, $q_1, q_2, a \in \{0, 1\}$, при $q_1q_2 \neq 01$, $\varphi(01, 0) = 10, \varphi(01, 1) = q_T$, $\varphi(q_T, a) = q_T$, $a \in \{0, 1\}$.

По определению $f(q_T) = 0$. Состояние $00 = 0a_1$, т.е. $k = 1$, следовательно $f(00) = \neg a_2 = 0$. Состояние $01 = a_1a_2$, т.е. $k = 2$, следовательно $f(01) = \neg a_3 = 1$. Состояние $10 = 1a_1$, т.е. $k = 1$, следовательно

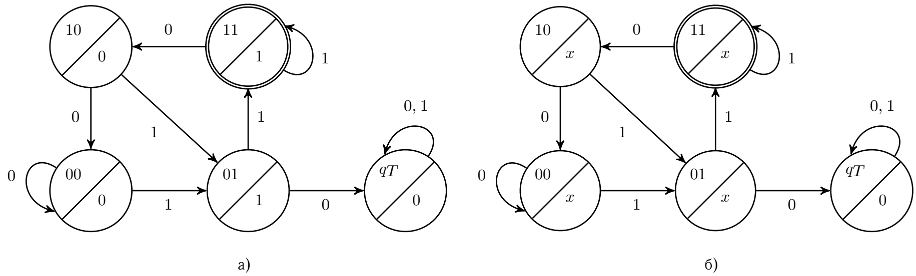

Рис. 5: а) Диаграмма Мура второго рода автомата $V(010)$; б) Традиционная диаграмма Мура автомата $V(010)$.

$f(10) = \neg a_2 = 0$. Префикс состояния 11 не совпадает ни с каким префиксом слова $a_1a_2a_3 = 010$, т.е. $k = 0$, следовательно $f(11) = \neg a_1 = 1$, и состояние 11 назначаем начальным.

Диаграмма Мура второго рода получившегося автомата приведена на рисунке 5 а), откуда легко получается традиционная диаграмма Мура этого автомата, изображенная на рисунке 5 б).

Убедимся, что если действовать по алгоритму, описанному в доказательстве леммы 2, мы получим такой же результат.

Функция выхода в тупиковом состоянии может принимать любое значение, пусть $f(q_T) = 0$.

$Q(1) = 01, \varphi(01, 0) = q_T$, поэтому $f(01) = 1$.

$Q(2) = \{00, 10\}$, $\varphi(00, 1) = \varphi(10, 1) = 01$, следовательно, пусть $f(00) = f(10) = 0$.

$Q(3) = \{11\}$, $\varphi(11, 0) = 10$, $f(11) = 1$.

Получили такую же функцию выхода.

Список литературы

[1] Трахтенброт Б.А., Барздинь Я.М. Конечные автоматы (поведение и синтез). — М.: Наука, 1970.

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

[3] Chalup S.K., Blair. Incremental training of first order recurrent neural networks to predict a context-sensitive language. // Neural Networks. - 2003. - Vol. 16, Iss. 7. - Pp.955 - 971.

[4] 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.

[5] 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. 482—487.

[6] Алешин С.В. Полугруппы и группы автоматов // Интеллектуальные системы. — 2013. — Т. 17, вып. 1–4. — С. 129–141.

[7] Александров Д.Е. Эффективные методы реализации проверки содержания сетевых пакетов регулярными выражениями // Интеллектуальные системы. — 2014. — Т. 18, вып. 1. — С. 37–60.

[8] Титова Е.Е. Конструирование движущихся изображений клеточными автоматами // Интеллектуальные системы. — 2014. — Т. 18, вып. 1. — С. 153–180.

[9] Бабин Д.Н. Частотные регулярные языки // Интеллектуальные системы. — 2014. — Т. 18, вып. 1. — С. 205–210.

[10] Иванов И.Е. О некоторых свойствах автоматов с магазинной памятью // Интеллектуальные системы. — 2014. — Т. 18, вып. 1. — С. 243–252.

[11] В.Б.Кудрявцев. Кафедра математической теории интеллектуальных систем (МаТИС) // Интеллектуальные системы. — 2014. — Т. 18, вып. 2. — С. 5–30.

[12] Часовских А.А. Условия полноты линейно-p-автоматных функций // Интеллектуальные системы. — 2014. — Т. 18, вып. 3. — С. 203–252.

[13] Александров Д.Е. Об оценках автоматной сложности распознавания классов регулярных языков // Интеллектуальные системы. — 2014. — Т. 18, вып. 4. — С. 161–190.

[14] Дементьев В.М. О звездной высоте регулярного языка и циклической сложности минимального автомата // Интеллектуальные системы. — 2014. — Т. 18, вып. 4. — С. 215–222.

[15] Кучеренко И.В. О минимизации монофункциональных классов бинарных клеточных автоматов с неразрешимым свойством обратимости. — 2014. — Т. 18, вып. 4. — С. 227–295.

[16] Якимец К.К. Об инвариантности характеристик конфигураций однородных структур. — 2014. — Т. 18, вып. 4. — С. 347–356.

[17] Иванов И.Е. О сохранении периодических последовательностей автоматами с магазинной памятью с однобуквенным магазином // Интеллектуальные системы. — 2015. — Т. 19, вып. 1. — С. 145–160.

[18] Летуновский А.А. Выразимость линейных автоматов относительно расширенной суперпозиции // Интеллектуальные системы. — 2015. — Т. 19, вып. 1. — С. 161–170.

[19] Гербус В.Г. О связи функций автомата и автоматной функции // Интеллектуальные системы. — 2015. — Т. 19, вып. 2. — С. 109–116.

[20] Миронов А.М. Критерий реализуемости функций на строках вероятностными автоматами Мура с числовым выходом // Интеллектуальные системы. — 2015. — Т. 19, вып. 2. — С. 149–160.

[21] Терехина И.Ю. Модель невлияния для квантовых автоматов // Интеллектуальные системы. — 2015. — Т. 19, вып. 2. — С. 183–190.

[22] Вереникин А.Г., Гасанов Э.Э. Об автоматной детерминизации множеств сверхслов // Дискретная математика. — 2006. Т. 18, № 2. — С. 84–97.

[23] Гасанов Э.Э. Прогнозирование периодических сверхсобытий автоматами // Интеллектуальные системы. — 2015. — Т. 19, вып. 1. — С. 23–34.

[24] Мастихина А.А. О частичном угадывании сверхслов // Интеллектуальные системы — 2007. — Т.11 вып.1-4 — С.609-619.

[25] Мастихина А.А. Критерий частичного предвосхищения общерегулярных свехсобытий // Дискретная математика. — 2011.—Т.23, вып.4 — С.103-114.

[26] Кудрявцев В.Б., Алешин С.В., Подколзин А.С. Введение в теорию автоматов. — М.: Наука, 1985.

Prediction of general regular superevents by automata

E. E. Gasanov , A. A. Mastikhina

A criterion for a prediction of general regular superevents and a criterion for a prediction of superiteration of regular events are proved in the article. Algorithms to construct predictive automata are presented.

Keywords: event prediction, automata, general regular superevents.