Записки научных
семинаров ПОМИ
Том 402, 2012 г.

Ю. И. Закс, Е. С. Скворцов

СИНХРОНИЗИРУЕМЫЕ СЛУЧАЙНЫЕ АВТОМАТЫ НАД 4-БУКВЕННЫМ АЛФАВИТОМ

§1. Введение и формулировка основного результата

Пусть $\mathcal{A} = (Q, \Sigma, \delta)$ – (*детерминированный конечный*) *автомат* с конечным множеством состояний Q, конечным входным алфавитом Σ и функцией перехода $\delta : Q \times \Sigma \to Q$, определяющей действие букв из алфавита Σ на состояния из Q. Действие букв естественным образом продолжается до действия слов над алфавитом Σ; результат действия слова w на состояние $\mathbf{q} \in Q$ обозначается через $\mathbf{q}w$:

$$\mathbf{q}w = \begin{cases} \mathbf{q}, & \text{если } w - \text{пустое слово,} \\ \delta(\mathbf{q}w', a), & \text{если } w = w'a \text{ для некоторых} \\ & \text{слова } w' \text{ и буквы } \ a \in \Sigma. \end{cases}$$

Слово w называется *синхронизирующим словом* автомата $\mathcal{A}$, если действие этого слова переводит все состояния этого автомата в одно состояние, т. е. $\mathbf{q}_1 w = \mathbf{q}_2 w$ для любых состояний $\mathbf{q}_1, \mathbf{q}_2 \in Q$. Автомат $\mathcal{A}$ называется *синхронизируемым*, если у него есть синхронизирующее слово.

Синхронизируемые автоматы активно применяются в различных областях: роботике, тестировании систем и протоколов, символической динамике и др. (см. обзоры [4, 6, 9]). С ними связан ряд интересных открытых вопросов, одним из которых является вопрос об оценке длины кратчайшего синхронизирующего слова в зависимости от числа состояний автомата. Наилучшая известная на текущий момент верхняя оценка длины кратчайшего синхронизирующего слова для автомата с n состояниями, равная $(n^3 - n)/6$, была получена Пэном [5] в 1983 г. Предположение о том, что эта длина не превосходит $(n-1)^2$, сформулированное Черни еще в середине 1960-х гг., доказано для некоторых специальных классов автоматов, но остается открытой проблемой в общем случае.

Ключевые слова: синхронизируемый автомат, случайный автомат, гипотеза Черни.

Первый автор получал поддержку РФФИ по гранту 10-01-00793.

На практике *медленно синхронизируемые автоматы*, т.е. автоматы с кратчайшим синхронизирующим словом длины $\Theta(n^2)$, встречаются исключительно редко[1]. С точки зрения практического применения синхронизируемых автоматов представляется важным изучить поведение длины кратчайшего синхронизирующего слова в среднем. Результаты вычислительных экспериментов (см., например, [7]) показывают, что оно существенно отличается от поведения этой величины в экстремальных случаях, которым традиционно уделяется основное внимание в исследованиях по синхронизируемым автоматам.

Далее в работе под случайной величиной будет пониматься дискретная случайная величина, в каждом случае это дополнительно оговариваться не будет.

Дадим строгое определение случайного автомата и сформулируем интересующие нас вопросы относительно его свойств, касающихся синхронизируемости.

Рассмотрим множество состояний Q и алфавит Σ. Выберем функцию перехода δ равномерно случайно из множества всюду определенных функций $\{\delta : Q \times \Sigma \to Q\}$. Получившаяся тройка (Q, Σ, δ) определяет *случайный конечный детерминированный автомат*. Следует отметить, что случайный автомат может быть построен следующим образом: для каждого состояния $q \in Q$ и для каждой буквы $a \in \Sigma$ выбираем $q' = \delta(q, a)$ равномерно случайно из Q. Под выбором "равномерно случайно" мы понимаем выбор, при котором каждый объект может быть выбран равновероятно.

Мы ставим следующие вопросы:

(1) Какой размер входного алфавита достаточен, чтобы почти все автоматы над алфавитом этого размера были синхронизируемы, и какой будет наиболее вероятная длина кратчайшего синхронизирующего слова для таких автоматов? (Под "почти всеми автоматами" мы понимаем долю автоматов, стремящуюся к 1 при $n \to \infty$. Утверждение, выполняющееся для почти всех объектов мы будем также называть выполняющимся "с высокой вероятностью".)

[1]В течение длительного времени единственной бесконечной серией таких автоматов была серия, построенная Черни [3]. Первые серии медленно синхронизируемых автоматов, существенно отличающиеся от серии Черни, были получены сравнительно недавно [1, 2].

(2) Какой размер входного алфавита достаточен, чтобы почти все автоматы над алфавитом этого размера были синхронизируемы и удовлетворяли гипотезе Черни?
(3) Какой размер входного алфавита достаточен, чтобы автомат над алфавитом этого размера был синхронизируем с конечной вероятностью? ("Конечной" мы называем вероятность, ограниченную снизу некоторой положительной константой при $n \to \infty$.)

В [8] мы дали частичные ответы на первые два вопроса для автоматов с n состояниями и $m(n)$ буквами (число букв зависит от числа состояний). В данной работе мы обращаемся к третьему вопросу и показываем, что случайный автомат с размером алфавита, не зависящим от числа состояний, синхронизируем с конечной вероятностью. Нашим основным результатом является следующая теорема.

Теорема 1. *Существует константа $p_0 > 0$ такая, что для любого натурального числа n случайный автомат $\mathcal{A} = (Q, \Sigma, \delta)$ с $|Q| = n$ и $|\Sigma| = 4$ синхронизируем с вероятностью большей, чем p_0.*

Отметим, что экспериментальные результаты подсказывают, что в действительности справедливо намного более сильное утверждение: случайный автомат $\mathcal{A} = (Q, \Sigma, \delta)$ с $|Q| = n$ и $|\Sigma| \geq 2$ синхронизируем с вероятностью, стремящейся к 1 при $n \to \infty$. Однако на данный момент теорема 1 – это максимум того, что нам удается доказать.

§2. Доказательство основного результата

Для начала докажем следующую техническую лемму.

Лемма 1. *Пусть X – случайная величина, принимающая значения из отрезка $[0, 1]$. Тогда имеет место неравенство*

$$\mathbf{P}\left(X \geq \mathbf{E}\left(X\right)/2\right) \geq \frac{\mathbf{E}\left(X\right)}{2 - \mathbf{E}\left(X\right)}.$$

Доказательство. Пусть c – константа такая, что $0 < c < 1$. В определении математического ожидания случайной величины X оценим сверху константой c значения X, не превышающие c, и константой 1 – значения, превышающие c. Получим следующее неравенство

$$\mathbf{E}\left(X\right) \leq c\mathbf{P}\left(X \leq c\right) + \left(1 - \mathbf{P}\left(X \leq c\right)\right),$$

или

$$\mathbf{P}\left(X \leq c\right) \leq \frac{1 - \mathbf{E}\left(X\right)}{1 - c},$$

или

$$\mathbf{P}\left(X \geq c\right) \geq \frac{\mathbf{E}\left(X\right) - c}{1 - c}.$$

Положив c равным $\mathbf{E}\left(X\right)/2$, завершаем доказательство. □

Пусть $\mathbf{u}, \mathbf{v}$ – пара состояний случайного автомата $\mathcal{A} = (Q, \Sigma, \delta)$. Определим для этой пары состояний процесс Roomba, цель которого – найти слово $w = a_1 \cdots a_k$ такое, что $\mathbf{u}w = \mathbf{v}w$.

На первом шаге процесса мы случайным образом выбираем букву $a_1 \in \Sigma$ и совершаем переход из $\mathbf{u}_0 = \mathbf{u}$ в $\mathbf{u}_1 = \mathbf{u}_0 a_1$ и из $\mathbf{v}_0 = \mathbf{v}$ в $\mathbf{v}_1 = \mathbf{v}_0 a_1$. Если $\mathbf{u}_1 = \mathbf{v}_1$, то процесс успешно завершается построением слова $w = a_1$, иначе он продолжается.

На m-м шаге процесса мы оказываемся в состояниях $\mathbf{u}_{m-1}$ и $\mathbf{v}_{m-1}$. (Не исключено, что какое-то из них или их оба мы уже посещали на предыдущих шагах процесса.) Выбираем букву a_m, которая ранее не применялась для переходов из состояния $\mathbf{u}_{m-1}$ или из состояния $\mathbf{v}_{m-1}$. Если мы смогли выбрать такую букву, то переходим с ее помощью из состояний $\mathbf{u}_{m-1}$, $\mathbf{v}_{m-1}$ в состояния $\mathbf{u}_m = \mathbf{u}_{m-1}a_m$, $\mathbf{v}_m = \mathbf{v}_{m-1}a$ соответственно, аналогично первому шагу. Назовем этот переход *ключевым переходом* процесса. Если $\mathbf{u}_m = \mathbf{v}_m$, то процесс завершается построением слова $w = a_1 a_2 \ldots a_m$, иначе он продолжается.

Если мы не можем выбрать букву (это означает, что к каждому из состояний $\mathbf{u}_{m-1}$, $\mathbf{v}_{m-1}$ мы уже применили все имеющиеся буквы алфавита), то действуем следующим образом. Запускаем поиск в ширину состояния, из которого не применена хотя бы одна буква алфавита, одновременно из состояний $\mathbf{u}_{m-1}$ и $\mathbf{v}_{m-1}$. Если поиск не находит такого состояния, то процесс заканчивается неуспехом. В противном случае мы находим состояние, достижимое, допустим, из состояния $\mathbf{u}_{m-1}$ по слову $z \in \Sigma^*$. Перейдем из $\mathbf{u}_{m-1}$ и $\mathbf{v}_{m-1}$ в $\mathbf{u}_m = \mathbf{u}_{m-1}z$ и $\mathbf{v}_m = \mathbf{v}_{m-1}z$ и продолжим процесс. Назовем эту часть процесса *поиском слова* z.

Отметим, что процесс Roomba схож с процессом Vacuum из [8], различается в них только принцип определения буквы или слова для следующего хода. Формальное описание процесса Roomba приведено на рис. 1.

Вход: Случайный автомат $\mathcal{A} = (Q, \Sigma, \delta)$ и пара состояний $\mathbf{u} \in Q, \mathbf{v} \in Q$
Выход: ***неуспех*** или слово $w = a_1 \dots a_k$ такое, что $\mathbf{u}w = \mathbf{v}w$
Описание процесса:
пусть $\Delta_q \subseteq \Sigma$, $w \in \Sigma^*$
Установить $\Delta_q = \varnothing$ для всех $\mathbf{q} \in Q$, $w = \varepsilon$
пока $\mathbf{u}w \neq \mathbf{v}w$
если $\Delta_{\mathbf{u}w} \cap \Delta_{\mathbf{v}w} \neq \Sigma$, **то** *ключевой переход*
Выбрать $a \in \Sigma \setminus (\Delta_{\mathbf{u}w} \cap \Delta_{\mathbf{v}w})$
Установить $\Delta_{\mathbf{u}w} = \Delta_{\mathbf{u}w} \cup \{a\}$, $\Delta_{\mathbf{v}w} = \Delta_{\mathbf{v}w} \cup \{a\}$
Установить $w = wa$
иначе
поиск слова z
Установить $w = wz$
Вернуть w

Рис. 1. Процесс Roomba.

Отметим ряд полезных свойств описанного процесса, совершаемого с автоматом над двухбуквенным алфавитом.

Предложение 1. *Пусть $\mathcal{A} = (Q, \Sigma, \delta)$ – случайный автомат такой, что $|Q| = n$ и $|\Sigma| = 2$. Тогда процесс* Roomba*, начав с любой пары состояний $\mathbf{u}, \mathbf{v} \in Q$, завершается построением слова w после ключевого перехода с вероятностью $1/n$.*

Доказательство. По определению ключевого перехода мы совершаем его по букве a_i, которая ранее не использовалась из состояния $\mathbf{u}_{i-1}$ или из состояния $\mathbf{v}_{i-1}$, допустим, из $\mathbf{u}_{i-1}$. Значит, мы выбираем состояние $\mathbf{u}_i$ равномерно случайно из Q и оно совпадет с некоторым состоянием $\mathbf{v}_i$ с вероятностью $1/n$. □

Предложение 2. *Существует константа $c_1 > 0$ такая, что для произвольного случайного автомата $\mathcal{A} = (Q, \Sigma, \delta)$ с $|Q| = n$ и $|\Sigma| = 2$ процесс* Roomba*, начав с любой пары состояний $\mathbf{u}, \mathbf{v} \in Q$, пройдет через ключевой переход по крайней мере $c_1 n$ раз с высокой вероятностью, если не завершится построением слова w ранее.*

Доказательство. Сначала покажем, что произвольное множество состояний из Q, размером меньшее чем n/e, с высокой вероятностью имеет исходящее ребро. Для фиксированного множества состояний

размера m, где $m < n/e$, вероятность того, что из него не выходит ни одного ребра, равна $(m/n)^{2m}$. Всего таких множеств размера m имеется $\binom{n}{m} \leq \left(\frac{ne}{m}\right)^m$. Из неравенства Буля (*неравенством Буля* для краткости мы будем называть тривиальную оценку вероятности объединения событий суммой вероятностей) получаем, что вероятность того, что существует множество размера m без исходящих ребер, меньше

$$\left(\frac{m}{n}\right)^{2m} \left(\frac{ne}{m}\right)^m = \left(\frac{me}{n}\right)^m \underset{n\to\infty}{\longrightarrow} 0.$$

Суммируя по всем таким m, получим, что с высокой вероятностью все множества размера менее $c_1 n$ для некоторой константы c_1 с $0 < c_1 < 1/e$ имеют исходящее ребро. Откуда легко выводится утверждение, что с высокой вероятностью из любого состояния $\mathbf{q} \in Q$ достижимо не менее $c_1 n$ состояний. При этом константа c_1 не зависит ни от вида автомата, ни от числа его состояний.

Таким образом, поиск в ширину слова z на некотором шаге процесса Roomba сделает по крайней мере $c_1 n$ шагов, если успешно не завершится ранее. В силу конечности всех объектов, для процесса в целом есть две возможности: сделать в какой-то момент указанное число шагов поиска и построить слово w ранее.

Заметим, что по определению поиска на путях из $\mathbf{u}_{m-1}$ и $\mathbf{v}_{m-1}$, помеченных словом z, встречаются только те ребра, которые были использованы нами ранее. Тот факт, что мы встречаем незнакомое ребро, означает, что мы могли завершить поиск с более коротким словом z. Таким образом, мы используем ребро в первый раз только при ключевом переходе.

Тот факт, что поиск из некоторого $\mathbf{u}_j$ сделал $c_1 n$ шагов, означает, что в автомате есть $2c_1 n$ уже просмотренных ребер. Все эти ребра были когда-то использованы в первый раз, то есть ключевой переход был произведен по меньшей мере $c_1 n$ раз. □

Установленные свойства позволяют доказать следующую лемму.

Лемма 2. *Существуют константы $p_0 > 0$ и $c_0 > 0$ такие, что для любого натурального числа n и любого случайного автомата $\mathcal{A} = (Q, \Sigma, \delta)$ такого, что $|Q| = n$ и $\Sigma = \{a, b\}$, вероятность события*

$$\frac{|\{(\mathbf{u}, \mathbf{v}) \in Q^2 \mid \exists w \ \mathbf{u}w = \mathbf{v}w\}|}{n^2} > c_0$$

превышает p_0. Иными словами, конечная доля пар состояний в случайном автомате синхронизируема с конечной вероятностью.

Доказательство. Возьмем пару состояний $(\mathbf{u}, \mathbf{v}) \in Q \times Q$ и попробуем синхронизировать ее при помощи процесса Roomba. Согласно предложению 2 ключевой переход будет произведен не менее $c_1 n$ раз для некоторой константы c_1, если процесс не завершится успехом ранее. В соответствии с предложением 1 при каждом ключевом переходе синхронизация произойдет с вероятностью $1/n$. Ключевые переходы независимы, так что используя неравенство Буля по всем ключевым переходам, мы получим, что пара состояний синхронизируема с вероятностью, ограниченной снизу некоторой константой c_2.

Следовательно, математическое ожидание случайной величины

$$|\{(\mathbf{u}, \mathbf{v}) \in Q \times Q \mid \exists w \ \mathbf{u}w = \mathbf{v}w\}|$$

больше $c_2 n^2$. Применение леммы 1 к случайной величине

$$X = \frac{|\{(\mathbf{u}, \mathbf{v}) \in Q \times Q \mid \exists w \ \mathbf{u}w = \mathbf{v}w\}|}{n^2}$$

завершает доказательство. □

Доказательство теоремы 1. Пусть $\Sigma = \{a, b, d, f\}$. По лемме 2 с конечной вероятностью p_0 существует подмножество $T \subset Q \times Q$ такое, что $|T| > c_0 n^2$ и любая пара состояний из T синхронизируема некоторым словом w_1 над алфавитом $\{a, b\}$. Покажем, что для любой пары состояний, не принадлежащей T, существует путь из нее в пару состояний из T.

Рассмотрим произвольную пару состояний $\mathbf{u}, \mathbf{v} \in Q \setminus T$. При доказательстве предложения 2 установлено, что из выделенного состояния с высокой вероятностью достижимы не менее $c_3 n$ состояний для некоторой константы c_3. Воспользовавшись этим фактом, получим, что из состояния $\mathbf{u}$ достижимо с помощью слов над алфавитом $\{d, f\}$ не менее $c_3 n$ состояний для некоторой константы c_3. Обойдем эти состояния поиском в ширину, параллельно выполняя те же переходы из состояния $\mathbf{v}$. В результате получим $c_3 n$ пар состояний, среди которых не менее $\frac{c_3 n}{2}$ различных. Вероятность того, что множество из $\frac{c_3 n}{2}$ случайных пар не пересекается с T, ограничено сверху выражением $(1 - c_0)^{c_3 n/2}$, которое стремится к 0 при $n \to \infty$. Обозначим через w_2 слово над алфавитом $\{d, f\}$, помечающее путь из $\mathbf{u}, \mathbf{v}$ в пару из T.

Применив неравенство Буля по всем парам состояний, получаем, что с конечной вероятностью каждая пара состояний автомата будет синхронизирована по слову w_1, либо по слову $w_2 w_1$. Как хорошо

известно [3], если любая пара пара состояний автомата может быть синхронизирована, то и автомат в целом синхронизируем. □

Благодарности. Авторы благодарят анонимного рецензента за ценные замечания к работе.

Литература

1. D. S. Ananichev, V. V. Gusev, M. V. Volkov, *Slowly synchronizing automata and digraphs.* — Math. Found. Comput. Sci., Lect. Notes Comput. Sci. **6281** (2010), 55–64.
2. D. S. Ananichev, M. V. Volkov, Yu. I. Zaks, *Synchronizing automata with a letter of deficiency* 2. — Theor. Comput. Sci. **376** (2007), 30–41.
3. J. Černý, *Poznámka k homogénnym eksperimentom s konečnými automatami.* — Matematicko-fyzikalny Časopis Slovensk. Akad. Vied **14**, No. 3 (1964), 208–216 (in Slovak).
4. A. Mateescu, A. Saloma, *Many-valued truth functions, Černý's conjecture and road-coloring.* — Bull. EATCS **68** (1999), 134–150.
5. J.-E. Pin, *On two combinatorial problems arising from automata theory.* — Ann. Discr. Math. **17** (1983), 535–548.
6. S. Sandberg, *Homing and synchronizing sequences, Model-Based Testing of Reactive Systems.* — Lect. Notes Comput. Sci. **3472** (2005), 5–33.
7. E. Skvortsov, E. Tipikin, *Experimental study of the shortest reset word of random automata.* — Implement. Appl. Automata, Lect. Notes Comput. Sci. **6807** (2011), 290–298.
8. E. Skvortsov, Yu. Zaks, *Synchronizing random automata.* — Discr. Math. Theor. Comput. Sci. **12**, No. 4 (2010), 95–108.
9. M. V. Volkov, *Synchronizing automata and the Černý conjecture.* — Languages Automata: Theory and Appl., Lect. Notes Comput. Sci. **5196**, (2008), 11–27.

Zaks Yu. I., Skvortsov E. S. Synchronizing random automata on 4-letter alphabet.

The paper deals with the synchronization of a random automaton that is sampled uniformly at random from the set of all automata with n states and m letters. We show that for $m = 4$ the probability that a random automaton is synchronizing is larger than a positive constant.

Институт математики и
компьютерных наук,
Уральский федеральный университет,
620083, Ленина 51, Екатеринбург Россия
E-mail: yuzaks@gmail.com,
skvortsoves@googlemail.com

Поступило 25 августа 2012 г.