О логических теориях и конечных автоматах на бесконечных последовательностях

Юрий Притыкин*

April 24, 2009

Аннотация

В работе показан пример связи логики и комбинаторики. Обсуждается связь вопросов разрешимости монадических теорий второго порядка бесконечных последовательностей и конечно-автоматных преобразователей, действующих на бесконечных последовательностях.

Abstract

In the paper an example of a connection between logic and combinatorics is shown. We discuss connections between decidability problems of monadic second-order theories over infinite words and finite transductions of infinite words.

1 Введение

Рассмотрим следующий общий вопрос: дана функция — какие у неё свойства и как научиться их определять? Ограничимся функциями вида $\mathbb{N} \to \mathbb{N}$, принимающими лишь конечное множество различных значений. Такие функции мы будем отождествлять с последовательностями букв конечного алфавита.

Рассмотрим следующие простые вопросы о последовательности x: входит ли в x символ a? Входит ли слово u? Входит ли слово u бесконечно много раз? Можно сформулировать и более сложные вопросы про последовательность. Различные, в том числе и перечисленные только что, свойства бесконечных последовательностей могут быть выражены в *логической теории первого порядка*. Под такой теорией для последовательности $x \in A^{\mathbb{N}}$ мы будем понимать следующее. Формально, в качестве структуры возьмём $\langle \mathbb{N}, S, <, X \rangle$, где $\mathbb{N}$ — множество натуральных чисел, которое пробегают индивидные переменные, S — двухместный предикат следования, $<$ — двухместный предикат порядка на натуральных числах, X — функциональный символ, интерпретируемый как последовательность $x\colon \mathbb{N} \to A$. В качестве теории берём обычную теорию

*Московский Государственный Университет им. М. В. Ломоносова, e-mail: yura@mccme.ru. Работа частично поддержана грантами РФФИ N 06-01-00122, 05-01-02803, а также грантом поддержки научных школ НШ-5351.2006.1 и Фондом поддержки молодых учёных “Конкурс Мёбиуса”.

первого порядка, истинность формул интерпретируем естественным образом. Во всех рассматриваемых нами теориях мы подразумеваем наличие двухместного предиката равенства, интерпретируемого естественным образом.

Ясно, что у такой реализации есть много вариантов, эквивалентных между собой по выразительным способностям. Например, можно было вместо двухместного предиката следования взять в структуру одноместную функцию следования. Можно было и не брать предикат следования вообще, потому что он выразим через неравенство. Точно так же вместо одной функции X можно было взять семейство предикатов X_a по одному для каждого $a \in A$, истинных ровно там, где в x стоит буква a. Можно обойтись и меньшим — логарифмическим по отношению к размеру алфавита последовательности — количеством предикатов. Кроме того, поскольку константа 0 выразима в определённой выше структуре, ясно, что можно было, не меняя множество выразимых формул, добавить в структуру все константы.

Теорию, определённую выше, будем обозначать $\mathrm{T}\langle x\rangle$.

Несложно видеть, что простые свойства последовательности x, типа упомянутых в начале раздела, можно выразить в теории $\mathrm{T}\langle x\rangle$. Например, формула $\forall p \exists q \exists r\,(q > p) \wedge S(q,r) \wedge X(q) = 0 \wedge X(r) = 1$ означает свойство вхождения в x бесконечное количество раз слова 01. Тем не менее, некоторые достаточно просто формулируемые свойства, которые, возможно, хотелось бы выразить, в теории первого порядка выразить нельзя, например, "бесконечное слово в алфавите $\{a, b\}$, в котором между любыми последовательными вхождениями b (такими, между которыми нет других вхождений b) входит нечётное количество букв a".

Гораздо больше можно выразить в более сильной *монадической теории второго порядка*. В такой теории, кроме обычных переменных по натуральным числам $p, q, \ldots$, разрешены также монадические переменные по множествам (или по одноместным предикатам) $P, Q, \ldots$. Вводятся также соответствующие атомарные формулы вида $P(p), Q(p), \ldots$, обозначающие "p принадлежит P", "p принадлежит Q"… Разрешены также кванторы по монадическим переменным. Такую теорию мы будем обозначать $\mathrm{MT}\langle x\rangle$.

Теории, аналогичные вышеперечисленным, но не расширенные последовательностью, будем обозначать соответственно $\mathrm{T}\langle \mathbb{N}, <\rangle$, $\mathrm{MT}\langle \mathbb{N}, <\rangle$.

Теория $\mathrm{MT}\langle x\rangle$ богаче теории $\mathrm{T}\langle x\rangle$. Например, упомянутое выше не выразимое в теории первого порядка свойство в монадической теории выражается так: $\forall p \forall q (X(p) \wedge X(q) \wedge (p < q) \wedge \forall r((p < r) \wedge (r < q) \to \neg X(r)) \to \exists Y(\forall u \forall v(S(u,v) \to (Y(u) \leftrightarrow Y(v))) \wedge Y(p) \wedge Y(q)))$.

Как и в случае с теориями первого порядка, в описанной формализации монадической теории многое можно реализовать по-другому. Например, можно отказаться от неравенства, потому что в монадической теории оно выразимо через следование. Мы будем переходить от одной реализации к другой для удобства.

Для формулы ϕ в любом из вышеописанных языков будем обозначать через $L(\phi)$ множество всех последовательностей x, для которых эта формула верна (то есть верна, если интерпретировать входящую в неё единственную свободную переменную по функциям X как x).

Естественным (и основным для нас) вопросом о любой теории является вопрос о её разрешимости. Теория *разрешима*, если существует алгоритм, который по любой замкнутой формуле определяет её истинность.

Основной вопрос здесь — полностью описать класс последовательностей, для которых разрешима их монадическая теория. В такой формулировке вопрос не решён, и, видимо, удовлетворительно решён быть не может.

Однако благодаря применению теории автоматов для достаточно широких классов последовательностей можно получить критерий разрешимости их монадических теорий. В разделе 2 мы определяем автоматы Бюхи и Мюллера на последовательностях и обсуждаем, как они связаны с вопросами разрешимости логических теорий. В разделе 3 мы определяем некоторые классы последовательностей со свойствами, близкими к свойствам периодических последовательностей. В разделе 4 мы обсуждаем результаты о конечно-автоматных преобразованиях последовательностей со свойствами типа почти периодичности и следствия из этих результатов о разрешимости монадических теорий таких последовательностей.

Подробнее о монадических теориях последовательностей см., например, [12, 18, 13]. Настоящая статья получена переработкой из нескольких разделов обзорной статьи [14]. В частности, в [14] можно найти доказательства многих результатов, которые ниже приводятся без доказательства. О морфических и автоматных последовательностях, о которых идёт речь в конце статьи, можно также прочесть в монографии [2].

2 Монадические теории и конечные автоматы

Назовём (недетерминированным) *автоматом Бюхи* совокупность $M = \langle A, Q, \tilde{q}, \Delta, F\rangle$, где A — входной алфавит, Q — множество состояний, $\tilde{q} \in Q$ — начальное состояние, $\Delta \subseteq Q \times A \times Q$ — множество переходов, $F \subseteq Q$ — множество заключительных состояний. Ходом автомата M на последовательности $x = x(0)x(1)x(2)\dots$ называется такая последовательность состояний $\rho = \rho(0)\rho(1)\rho(2)\dots$, что $\rho(0) = \tilde{q}$ и $\langle \rho(i), x(i), \rho(i+1)\rangle \in \Delta$ для любого i. Мы говорим, что автомат M *принимает* x, если существует хотя бы один ход ρ автомата M на x, для которого хотя бы одно состояние, встречающееся в ρ бесконечное количество раз, входит в множество заключительных состояний F. Определяя *детерминированный автомат Бюхи*, множество переходов $\Delta \subseteq Q \times A \times Q$ можно заменить на функцию переходов $\delta\colon Q \times A \to Q$ (с естественными изменениями для определения хода автомата).

Есть немного другой вариант понятия автомата на последовательности. Назовём (недетерминированным) *автоматом Мюллера* совокупность $M = \langle A, Q, \tilde{q}, \Delta, \mathcal{F}\rangle$, где A — входной алфавит, Q — множество состояний, $\tilde{q} \in Q$ — начальное состояние, $\Delta \subseteq Q \times A \times Q$ — множество переходов, $\mathcal{F} \subseteq 2^Q$ — множество заключительных макросостояний. Здесь под *макросостоянием* мы понимаем элемент множества 2^Q, то есть произвольное подмножество множества состояний Q. Ходом автомата M на последовательности $x = x(0)x(1)x(2)\dots$ называется такая последовательность состояний $\rho = \rho(0)\rho(1)\rho(2)\dots$, что $\rho(0) = \tilde{q}$ и $(\rho(i), x(i), \rho(i+1)) \in \Delta$ для любого i. Назовём *пределом (предельным макросостоянием)* автомата M на последовательности x вдоль хода ρ множество всех таких состояний, которые встречаются в ρ бесконечное количество раз, обозначим это множество через $\lim_\rho M$. Мы говорим, что автомат M *принимает* x, если существует хотя бы один ход ρ автомата M на x, для которого $\lim_\rho M \in \mathcal{F}$.

Другими словами, слово принимается, если хотя бы на каком-нибудь ходе предельное макросостояние принадлежит множеству заключительных макросостояний $\mathcal{F}$. Аналогично предыдущему, определяя *детерминированный автомат Мюллера*, множество переходов $\Delta \subseteq Q \times A \times Q$ можно заменить на функцию переходов $\delta\colon Q \times A \to Q$ (с естественными изменениями для определения хода автомата).

Для автомата M Бюхи или Мюллера множество всех последовательностей, которые принимаются автоматом M, обозначим $L(M)$.

Для примера рассмотрим множество L последовательностей в алфавите $\{a, b, c\}$, в которые если a входит бесконечное количество раз, то и b входит бесконечное количество раз. Автоматы Мюллера и Бюхи, принимающие в точности слова множества L, показаны на рис. 1 (автомат Бюхи понадобился недетерминированный).

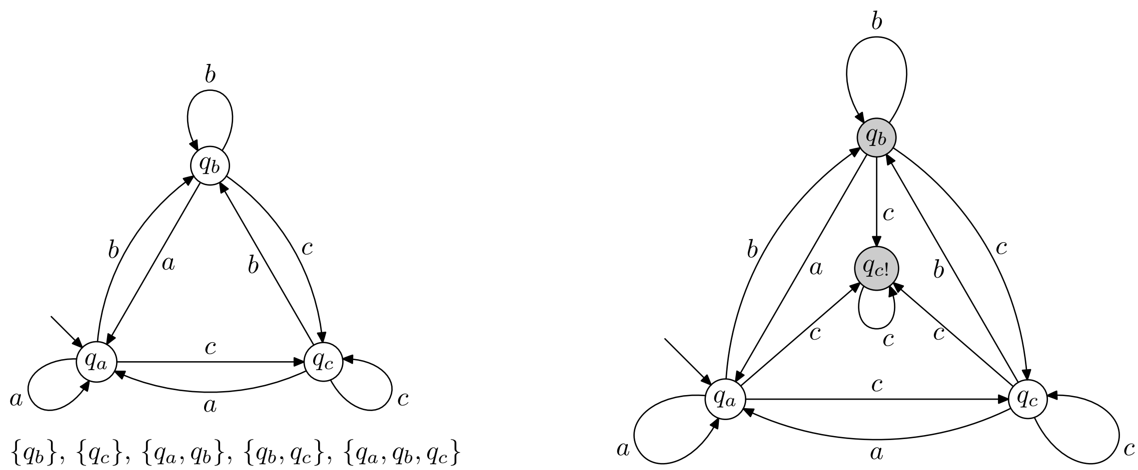

Рис. 1: Пример автомата Мюллера (слева) и автомата Бюхи (справа), принимающих одно и то же множество последовательностей. Множество принимающих макросостояний автомата Мюллера дано списком. Принимающие состояния автомата Бюхи отмечены тёмным цветом.

Это пример общей ситуации.

Теорема 1 ([7]). *Недетерминированные автоматы Бюхи, недетерминированные автоматы Мюллера и детерминированные автоматы Мюллера распознают один и тот же класс множеств последовательностей. Более того, по автомату одного типа можно получать эквивалентный автомат другого типа алгоритмически.*

Множество последовательностей, распознаваемое автоматом Мюллера или автоматом Бюхи, будем называть *регулярным*. Детерминированные автоматы Бюхи распознают меньший класс множеств.

Оказывается, что между описанием множеств последовательностей с помощью конечных автоматов и описанием их формулами монадического языка есть прямая связь.

Теорема 2 ([3]). *Существует алгоритм, который по каждому автомату Бюхи M строит формулу ϕ монадического языка, такую что $L(M) = L(\phi)$, и наоборот, по любой формуле ϕ строит такой автомат Бюхи M, что $L(\phi) = L(M)$.*

Следствие 3. *Множество последовательностей регулярно тогда и только тогда, когда оно выразимо в монадическом языке.*

Следствие 4 ([3]). *Теория* $\mathrm{MT}\langle \mathbb{N}, < \rangle$ *разрешима.*

Нас же интересует ситуация, когда теория $\mathrm{MT}\langle \mathbb{N}, < \rangle$ расширена последовательностью. Несложно видеть, что выполняется следующее следствие из теоремы Бюхи (теоремы 2).

Следствие 5. *Монадическая теория последовательности x разрешима тогда и только тогда, когда существует алгоритм, который по любому автомату Бюхи (или любому детерминированному автомату Мюллера) может определить, принимает ли этот автомат последовательность x или нет.*

Следствие 5 даёт способ устанавливать разрешимость монадических теорий некоторых типов последовательностей.

3 Почти периодичность

Последовательность x *периодическая*, если для некоторого $T \in \mathbb{N}$, называемого *периодом*, имеем $x(i) = x(i+T)$ для любого $i \in \mathbb{N}$. В соответствии с общепринятым соглашением, периодом мы называем также и слово $x[0, T-1]$. Будем называть последовательность *заключительно периодической*, если то же выполнено для всех i, начиная с некоторого K. Тогда $x[0, K-1]$ называется *предпериодом*. Предпериодом будем называть также и число K. Последовательность, не являющаяся заключительно периодической, называется *апериодической*. Множество всех периодических последовательностей обозначим $\mathcal{P}$, множество заключительно периодических обозначим $\mathcal{EP}$. Рассмотрим некоторые расширения этих классов.

Последовательность x называется *почти периодической*, если для каждого её фактора u найдётся такое натуральное l, что на каждом отрезке длины l последовательности x найдётся вхождение слова u. Тем самым, любое слово, входящее в почти периодическую последовательность, входит в неё бесконечное количество раз. Через $\mathcal{AP}$ (almost periodic) будем обозначать класс всех таких последовательностей. Ясно, что для проверки почти периодичности достаточно только убедиться в повторяемости с ограниченными интервалами всех префиксов, а не всех факторов. Почти периодические последовательности называют также *равномерно рекуррентными* и *минимальными*.

Будем называть последовательность x *заключительно почти периодической*, если некоторый её суффикс почти периодичен. Класс всех таких последовательностей обозначим $\mathcal{EAP}$ (eventually almost periodic).

Последовательность x называется *обобщённо почти периодической*, если для каждого её фактора u, входящего в неё бесконечное число раз, найдётся такое натуральное l, что на каждом отрезке длины l последовательности x найдётся вхождение слова u. Класс всех таких последовательностей обозначим через $\mathcal{GAP}$ (generalized almost periodic).

Если $x \in \mathcal{EAP}$, то минимальное такое n, что $x[n, \infty) \in \mathcal{AP}$, будем называть *минимальным префиксом* и обозначать $\mathrm{pr}(x)$. Заметим, что для любого $m \geqslant \mathrm{pr}(x)$ имеем $x[m, \infty) \in \mathcal{AP}$.

И наконец, назовём последовательность x *рекуррентной*, если каждое слово, которое в неё входит, обязательно входит бесконечное количество раз. Ясно, что если последовательность рекуррентная и обобщённо почти периодическая, то она почти периодическая. Класс рекуррентных последовательностей будем обозначать $\mathcal{R}$. Последовательность *заключительно рекуррентная*, если некоторый её суффикс рекуррентен. Класс таких последовательностей обозначим $\mathcal{ER}$.

Регулятором почти периодичности последовательности $x \in \mathcal{GAP}$ назовём функцию $\mathrm{R}_x\colon \mathbb{N} \to \mathbb{N}$, которая на числе n равна минимальному такому l, что каждое слово длины n, которое входит в x бесконечное количество раз, встретится на любом отрезке длины l последовательности x, а также любое слово длины n, которое входит в x конечное количество раз, не входит в $x[l, \infty)$ (второе важно только для обобщённо почти периодических последовательностей, не являющихся почти периодическими). Часто вместо регулятора нам будет достаточно рассматривать только какую-то верхнюю оценку на него, то есть функцию f, такую что $f(n) \geqslant \mathrm{R}_x(n)$ для всех n — в этом случае мы пишем $f \geqslant \mathrm{R}_x$. Если говорить более точно, регулятор почти периодичности объединяет в себе две функции: одна следит за расстояниями между вхождениями слов, входящих бесконечное количество раз, а другая следит за тем префиксом, которым ограничивается вхождение слов, входящих конечное количество раз. Иногда стоит эти функции разделять явно, хотя на протяжении настоящей статьи мы этого делать не будем.

Несложно видеть, что $\mathcal{P} \subset \mathcal{AP} \subset \mathcal{EAP} \subset \mathcal{GAP}$. Оказывается, все эти включения строгие. Например, известная последовательность Туэ–Морса

$$\mathbf{t} = 01101001100101101001011001101001\ldots$$

— пример последовательности из $\mathcal{AP}$, но не из $\mathcal{P}$. Другой известный пример — последовательность Фибоначчи

$$\mathbf{f} = 0100101001001010010100100101001001\ldots$$

Последовательность $\mathbf{t}$ получается, если взять подстановку $0 \to 01$, $1 \to 10$, и, начав с символа 0, применить к нему подстановку, потом применить ту же подстановку к результату, и т. д. (каждый раз подстановка применяется ко всем буквам слова одновременно). Каждое следующее слово начинается с предыдущего. В пределе получится бесконечная последовательность $\mathbf{t}$. Аналогично можно получить $\mathbf{f}$ с помощью подстановки $0 \to 01$, $1 \to 0$. Подробнее о таком способе построения последовательностей см., например, [2]. Последовательности $\mathbf{t}$ и $\mathbf{f}$ обладают большим количеством интересных свойств и часто встречаются в комбинаторике слов, например, см. [1, 2, 6].

Другое семейство примеров почти периодических последовательностей (идущее из символической динамики, где это понятие и появилось [8, 9]) получается следующим образом. Разделим окружность длины 1 на несколько дуг, каждой из которых сопоставим некоторый символ конечного алфавита. Отметим какую-нибудь точку на окружности и начнём делать шаги из этой точки по окружности с некоторым фиксированным иррациональным шагом. При этом будем записывать, в какую дугу попадаем. Таким образом запишется почти периодическая последовательность (мы, возможно, иногда будем попадать на границу между дугами, но такое произойдёт не более одного раза для каждой границы, поэтому можем начать записывать с того места, когда это уже не происходит).

Неравенство $\mathcal{AP} \subsetneq \mathcal{EAP}$ очевидно (можно взять в качестве примера 10000000...). Неравенство $\mathcal{EAP} \subsetneq \mathcal{GAP}$ было доказано в [15]. Кроме того, выполнены включения $\mathcal{P} \subset \mathcal{EP} \subset \mathcal{EAP}$, все из которых тоже, очевидно, строгие. Соотношения между введёнными классами проиллюстрированы на рис. 2.

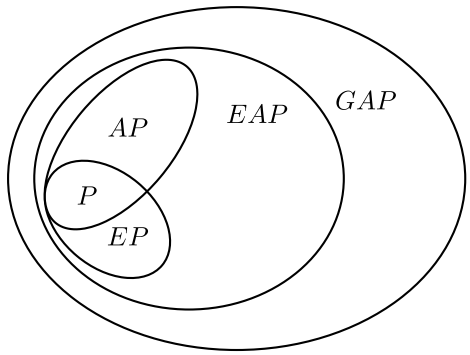

Рис. 2: Классы последовательностей.

Мы говорим, что последовательность эффективно обобщённо почти периодическая, если она вычислима, а также некоторая верхняя оценка на её регулятор почти периодичности вычислима. Класс таких последовательностей обозначим $\mathcal{GAP}^e$. Другими словами, для эффективно обобщённо почти периодической последовательности существует алгоритм, который, во-первых, по любому n может выдать n-й элемент последовательности, и во-вторых, для любого слова u может указать такое число l (не обязательно минимальное возможное), что если u входит в x бесконечное количество раз, то оно входит на любом отрезке длины l, а если u входит конечное количество раз, то оно не входит в суффикс $x[l, +\infty)$.

4 Конечно-автоматные преобразователи

Конечные автоматы, участвующие в определении конечно-автоматных преобразований бесконечных последовательностей, отличаются от автоматов Бюхи и Мюллера, определённых выше. Во-первых, у них есть начальное состояние, но нет принимающих. Во-вторых, на каждом переходе написана не только входная буква, но и выходная. Такой автомат читает по очереди буквы последовательности, записанные на входной ленте, и в соответствии с прочитанным меняет каждый раз текущее состояние, выдаёт букву на выходную ленту и переходит к следующей входной букве. Последовательность, полученная в итоге на выходной ленте, и есть результат конечно-автоматного преобразования.

Дадим теперь формальное определение. *Конечно-автоматным преобразователем* назовём совокупность $M = \langle A, B, Q, \tilde{q}, \lambda, \mu \rangle$, где A и B — конечные множества, называемые соответственно входной и выходной алфавит, Q — конечное множество состояний,

$\tilde{q} \in Q$ — выделенное состояние, называемое начальным, и

$$\lambda\colon Q \times A \to B^*, \qquad \mu\colon Q \times A \to Q$$

— функции переходов. Пусть $x \in A^{\mathbb{N}}$. Последовательность $(p_n)_{n=0}^{\infty}$ элементов множества Q назовём *ходом преобразователя M на x*, если $p_0 = \tilde{q}$ и для каждого n выполняется $p_{n+1} = \mu(p_n, x(n))$. Последовательность $M(x)$, определяемую как $M(x)(n) = \lambda(p_n, x(n))$, где $(p_n)_{n=0}^{\infty}$ — ход преобразователя M на x, назовём *образом последовательности x под действием M*. Определение иллюстрируется на рис. 3.

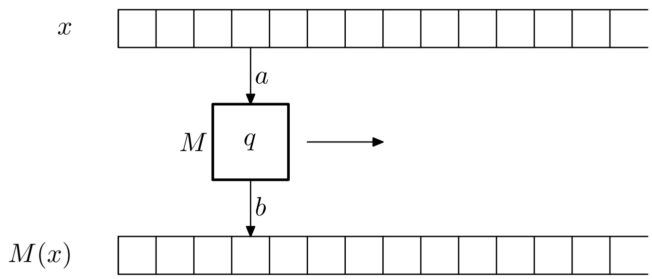

Рис. 3: Конечно-автоматное преобразование последовательности. Конечный автомат M, последовательность x, текущее состояние q, входная буква a, выходная буква b. Автомат идёт слева направо вдоль последовательности.

Если для каждых $a \in A$, $q \in Q$ выполнено $|\lambda(q, a)| = 1$, то преобразователь M называется *равномерным*. Применение произвольного конечно-автоматного преобразователя к последовательности можно представить как последовательное применение равномерного конечно-автоматного преобразователя и некоторого морфизма.

Предложение 6. *Пусть $M = \langle A, B, Q, \tilde{q}, \lambda, \mu\rangle$ — конечно-автоматный преобразователь, x — последовательность. Тогда существует такой равномерный конечно-автоматный преобразователь M' и такой морфизм ϕ, что $M(x) = \phi(M'(x))$.*

Доказательство. Действительно, положим $M' = \langle A, Q\times A, Q, \tilde{q}, \lambda, \mu'\rangle$, так что $\mu'(q, a) = \langle q, a\rangle$ для любых $q \in Q$ и $a \in A$. Определим также морфизм $\phi\colon Q \times A \to B$, так что $\phi(\langle q, a\rangle) = \mu(q, a)$ для любых $q \in Q$ и $a \in A$. Ясно тогда, что $M(x) = \phi(M'(x))$. □

Поэтому часто для упрощения ситуации мы ограничиваемся рассмотрением равномерных конечно-автоматных преобразователей. Если $[i, j]$ — вхождение слова u в последовательность x, причём $p_i = q$, где $(p_n)_{n=0}^{\infty}$ — ход преобразователя M на x, то будем говорить, что преобразователь M *подходит* к этому вхождению слова *u в состоянии q*.

По существу, следующая теорема доказана в [18] (см. также [10]). Мы приводим её с полным доказательством (по [10]).

Для произвольной функции g обозначим $\underbrace{g \circ g \circ \cdots \circ g}_{m}$ через g^m.

Теорема 7 ([18, 10]). *Пусть M — конечно-автоматный преобразователь с m состояниями, и $x \in \mathcal{GAP}$. Тогда верно следующее.*

1. $M(x) \in \mathcal{GAP}$.

2. *Пусть M — равномерный преобразователь. Тогда $\mathrm{R}_{M(x)}(n) \leqslant h(h(n))$ для всех n, где $h(n) = g^m(n) - 1$ и $g(n) = \mathrm{R}_x(n) + 1$.*

3. *Если $x \in \mathcal{GAP}^e$, то $M(x) \in \mathcal{GAP}^e$.*

Лемма 8. *Пусть M — равномерный конечно-автоматный преобразователь с m состояниями, и $x \in \mathcal{GAP}$. Пусть $v = M(x)[i, j]$ — вхождение слова длины n в $M(x)$, такое что $i \geqslant h(n)$, где $h(t) = g^m(t) - 1$, $g(t) = \mathrm{R}_x(t) + 1$. Тогда найдётся такое r, что $j - h(n) \leqslant r \leqslant i - 1$ и $M(x)[r, r + n - 1] = v$.*

Доказательство. Будем считать сначала, что $v = M(x)[i, j]$ — просто достаточно далёкое от начала вхождение слова v в $M(x)$. Объясним, как найти искомое $v = M(x)[r, s]$. При этом мы будем делать разные допущения, которые подытожим и сформулируем точно позднее.

Итак, пусть $v = M(x)[i, j]$ и $u_1 = x[i, j]$ — прообраз v в x. Пусть преобразователь M подходит к позиции i, находясь в состоянии q_1. Если i достаточно велико, найдётся вхождение $u_1 = x[i_2, j_2]$ левее $[i, j]$, но достаточно близко. Если M подходит к i_2 в состоянии q_1, то $M(x)[i_2, j_2] = v$. Иначе M подходит к i_2 в каком-то состоянии $q_2 \neq q_1$ (рис. 4).

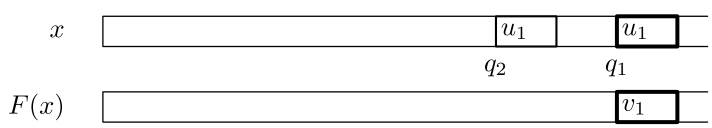

Рис. 4: Иллюстрация к доказательству теоремы 7.

Положим $u_2 = x[i_2, j]$. Если i_2 достаточно велико, то найдётся вхождение $u_2 = x[i_3, j_3]$ левее $[i_2, j]$, но достаточно близко. Если M подходит к i_3 в состоянии q_1, то $M(x)[i_3, i_3 + n - 1] = v$, так как u_2 начинается с u_1. Если M подходит к i_3 в состоянии q_2, то $M(x)[i_3, j_3] = M(x)[i_2, j]$, и тогда $M(x)[j_3 - n + 1, j_3] = v$, так как u_2 заканчивается словом u_1. В худшем случае M подходит к i_3 в состоянии q_3, таком что $q_3 \neq q_2$ и $q_3 \neq q_1$. При этом положим $u_3 = x[i_3, j]$ (рис. 5).

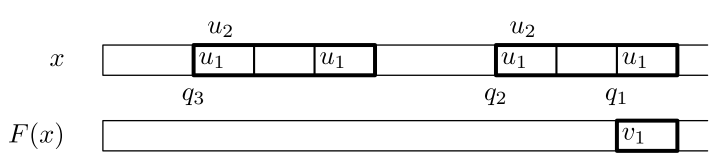

Рис. 5: Иллюстрация к доказательству теоремы 7.

Рассуждая так дальше, мы, всё время выбирая худший случай, найдём такие $i_2, \ldots, i_{m+1}$, что к каким-то двум из $i_1 = i, i_2, \ldots, i_{m+1}$ преобразователь M подходит в одинаковых состояниях. Таким образом, слово v обязательно встретится в $M(x)[i_{m+1}, j - 1]$.

Проанализируем теперь вышеприведённое рассуждение. Вначале мы ищем вхождение $u_1 = x[i_2, j_2]$. Это можно сделать, если $i \geqslant \mathrm{R}_x(n)$ — исходя из определения обобщённой почти периодичности, это означает, что u_1 входит в x бесконечно много раз.

При этом получится $(j-1)-i_2+1 \leqslant \mathrm{R}_x(n)$ — этого достаточно, чтобы на отрезке $x[i_2, j-1]$ нашлось вхождение u_1. Отсюда $i_2 \geqslant j - \mathrm{R}_x(n)$ и $|u_2| = j - i_2 + 1 \leqslant \mathrm{R}_x(n)+1$.

Далее аналогично получаем, что при $i_2 \geqslant \mathrm{R}_x(\mathrm{R}_x(n)+1) \geqslant \mathrm{R}_x(|u_2|)$ действительно можно найти вхождение $u_2 = x[i_3, j_3]$, причём $i_3 \geqslant j - \mathrm{R}_x(\mathrm{R}_x(n)+1)$ и $|u_3| \leqslant \mathrm{R}_x(\mathrm{R}_x(n)+1)+1$.

Положим $g = \mathrm{R}_x + 1$. Аналогично рассуждая, получим, что при $i_m \geqslant g^m(n)-1$ можно выбрать $i_{m+1} \geqslant j - g^m(n)+1$. □

Доказательство теоремы 7. Пусть $x \in \mathcal{GAP}$, и на x действует конечный преобразователь M с m состояниями.

1) Из предложения 6 следует, что достаточно доказать утверждение только для равномерного M, так как морфизмы сохраняют обобщённую почти периодичность. Для равномерного M из леммы 8 следует, что если слово v длины n входит в $M(x)$ бесконечно много раз, то оно входит на каждом отрезке длины $g^m(n)-1$ в $M(x)$. Поэтому $M(x)$ обобщённо почти периодическая.

2) Покажем, как найти оценку сверху на $\mathrm{R}_{M(x)}$. Обозначим $h(n) = g^m(n)-1$, где $g(n) = \mathrm{R}_x(n)+1$.

Пусть слово v длины n входит в $M(x)$ бесконечно много раз. Тогда по 1) оно встретится на любом отрезке длины $h(n)$.

Пусть теперь слово v длины n входит в $M(x)$ конечное количество раз. Докажем, что тогда v не входит в $M(x)$ правее позиции $h(h(n))$. Действительно, предположим, что это не так. Тогда на каждом отрезке длины $h(n)$ слова $M(x)[0, h(h(n))-1]$ найдётся вхождение v, так как по 1) от каждого такого вхождения можно найти ещё одно слева не дальше, чем на расстоянии $h(n)$. Но v входит в $M(x)$ лишь конечное количество раз. Поэтому в $M(x)$ найдётся слово w длины $h(n)$ (где-то сильно справа), которое не содержит v. Тогда всюду слева от него на любом отрезке длины $h(h(n))$ найдётся вхождение w, значит, и в $M(x)[0, h(h(n))-1]$ тоже. Противоречие, так как на любом отрезке длины $h(n)$ в $M(x)[0, h(h(n))-1]$ есть вхождение v, но w не содержит v и входит в $M(x)[0, h(h(n))-1]$.

Таким образом, объединяя последние два абзаца, получаем оценку $\mathrm{R}_{M(x)}(n) \leqslant h(h(n))$.

3) Пусть теперь последовательность x эффективно обобщённо почти периодическая, т. е. x вычислима и регулятор R_x вычислим. Аналогично п. 1), достаточно рассматривать только случай равномерного M, так как морфизмы сохраняют эффективную обобщённую почти периодичность. Ясно, что зная M и умея вычислять x, мы можем вычислять $M(x)$. П. 2) позволяет также вычислять оценку сверху на $\mathrm{R}_{M(x)}$. Значит, $M(x)$ эффективно обобщённо почти периодическая. □

В частности, мы видим, что для $x \in \mathcal{EAP}$ или $x \in \mathcal{AP}$ и произвольного конечно-автоматного преобразования M выполнено $M(x) \in \mathcal{GAP}$. Теорема 10 усиливает этот результат.

Имея и самостоятельный интерес, теорема 7 позволяет также получить следующее

Следствие 9. *Для $x \in \mathcal{GAP}$ теория $\mathrm{MT}\langle x\rangle$ разрешима тогда и только тогда, когда $x \in \mathcal{GAP}^e$.*

Доказательство. $\Rightarrow$. Пусть для $x \in \mathcal{GAP}$ теория $\mathrm{MT}\langle x\rangle$ разрешима. Тогда для каждого n и для каждого возможного символа a проверяем, не верно ли $x(n) = a$, и так перебором находим значение $x(n)$. Следовательно, x вычислима. Также перебором можно вычислять R_x, поскольку для любых n и l можно записать формулой тот факт, что $\mathrm{R}_x(n) \leqslant l$. Действительно, во-первых, мы можем найти перебором, какие элементы входят в последовательность (несмотря на то, что потенциальных претендентов, как мы договорились, может быть бесконечное количество), поскольку можно выразить формулой тот факт, что все символы, кроме символов из некоторого фиксированного конечного множества, в последовательности не встречаются. Во-вторых, слов фиксированной длины конечное число, поскольку последовательность конечно-значная.

$\Leftarrow$. Пусть теперь $x \in \mathcal{GAP}^e$.

По следствию 5 для разрешимости $\mathrm{MT}\langle x\rangle$ достаточно уметь по любому детерминированному автомату Мюллера, запущенному на x, определять, принимает ли он x или нет.

Пусть M — какой-нибудь детерминированный автомат Мюллера, действующий на x. Рассмотрим конечно-автоматный преобразователь M', полученный из M следующим образом: внутреннее устройство M' такое же, как и у M, при этом про принимающие макросостояния M мы забываем. На входных последовательностях M' работает точно так же, как и M, а на выходную ленту записывает на каждом шаге своё текущее состояние. По теореме 7 из $x \in \mathcal{GAP}^e$ следует $M'(x) \in \mathcal{GAP}^e$.

Для любой $x \in \mathcal{GAP}^e$ легко вычислить, какие символы входят в x бесконечно много раз: для этого достаточно посмотреть, какие символы входят в отрезок $x[f(1), 2f(1)-1]$, где f — какая-нибудь вычислимая оценка сверху на регулятор (а “посмотреть” мы можем, потому что сама последовательность x тоже вычислима).

Таким образом, мы можем найти множество всех таких состояний автомата M, которые встречаются бесконечное количество раз в процессе работы M на x. Значит, мы можем проверить, принимает ли автомат M последовательность x. □

В частности, из следствия 9 получаем, что монадические теории последовательностей Туэ–Морса $\mathbf{t}$ и Фибоначчи $\mathbf{f}$ разрешимы (для этих последовательностей несложно научиться оценивать сверху регулятор).

Следующий результат продолжает результат теоремы 7, но, видимо, не имеет приложений и следствий для логических теорий.

Теорема 10 ([15, 16]). *1) Множество заключительно почти периодических последовательностей замкнуто относительно конечно-автоматных преобразований.*

2) Пусть F — конечный автомат с n состояниями, $x \in \mathcal{AP}$. Тогда $F(x) \in \mathcal{EAP}$ и

$$\mathrm{pr}(F(x)) \leqslant \mathrm{R}_x^n(1) + \mathrm{R}_x^{n-1}(1) + \cdots + \mathrm{R}_x(1).$$

“Экспоненциальные” оценки (в смысле итераций функций) для регулятора почти периодичности в теореме 7 и для префикса в теореме 10 не могут быть существенно улучшены, как показано в [17, 11].

Одним из основных шагов в доказательстве теоремы 10 является следующий результат, представляющий и самостоятельный интерес.

Назовём конечно-автоматный преобразователь $M = \langle A, B, Q, \tilde{q}, \lambda, \mu\rangle$ *обратимым*, если для каждых $q \in Q$ и $a \in A$ существует ровно одно состояние $q' \in Q$, такое что $\mu(q', a) = q$. Другими словами, в таком преобразователе каждая буква входного алфавита осуществляет взаимно однозначное отображение множества состояний в себя. Находясь в некотором состоянии и зная последовательность предыдущих входных символов, можно восстановить и последовательность пройденных состояний (в этом и заключается свойство обратимости).

Предложение 11 ([16]). *Пусть M — обратимый равномерный конечно-автоматный преобразователь с m состояниями. Тогда если $x \in \mathcal{AP}$, то $M(x) \in \mathcal{AP}$.*

Ещё один результат из этой серии утверждает замкнутость класса заключительно рекуррентных последовательностей.

Теорема 12 ([14]). *Пусть M — конечно-автоматный преобразователь. Тогда если $x \in \mathcal{ER}$, то $M(x) \in \mathcal{ER}$.*

Доказательства теорем 10, 12 и предложения 11 можно найти в [14].

Приведём ещё один аналогичный результат. Способ, которым мы определили последовательности Туэ–Морса и Фибоначчи в разделе 3, можно обобщить. Назовём *подстановкой*, или *морфизмом*, любое отображение вида $\phi\colon A^* \to A^*$ (здесь A^* — множество всех конечных слов над алфавитом A), такое что для любых слов $u, v \in A^*$ выполнено $\phi(uv) = \phi(u)\phi(v)$. Ясно, что тогда морфизм полностью определяется своими значениями на однобуквенных словах.

Пусть $\phi(s) = su$ для некоторых $s \in A$, $u \in A^*$. Тогда для всех натуральных $m < n$ слово $\phi^n(s)$ начинается со слова $\phi^m(s)$, так что можно корректно определить $\phi^\infty(s) = \lim_{n\to\infty} \phi^n(s) = su\phi(u)\phi^2(u)\phi^3(u)\ldots$ Если для всякого n выполнено $\phi^n(u) \neq \Lambda$, то последовательность $\phi^\infty(s)$ бесконечна. Последовательности вида $\phi^\infty(s)$ называются *чисто морфическими.* Чисто морфические последовательности являются неподвижными точками морфизма: для них выполнено $\phi(\phi^\infty(s)) = \phi^\infty(s)$ (если продолжить ϕ на бесконечные последовательности естественным образом). Например, последовательности Туэ–Морса и Фибоначчи — чисто морфические.

Если в чисто морфической последовательности некоторые буквы отождествить, получим морфическую последовательность. Более формально, морфизм $h\colon A^* \to A^*$, переводящий однобуквенные слова в однобуквенные слова, называется *кодированием.* Последовательности вида $h(\phi^\infty(s))$, где $\phi^\infty(s)$ — чисто морфическая последовательность, h — кодирование, называются *морфическими.*

Теорема 13 ([5]). *Множество морфических последовательностей замкнуто относительно конечно-автоматных преобразований.*

Доказательство теоремы 13 можно найти также в [2].

Следствие 14. *Для любой морфической последовательности x теория* $\mathrm{MT}\langle x\rangle$ *разрешима.*

Действительно, можно следовать общей схеме из доказательства следствия 9. Проанализировав доказательство теоремы 13, можно увидеть, что она позволяет по морфической последовательности (которую можно задать конечным набором конечных объектов — алфавит, морфизм, начальная буква, кодирование) и конечно-автоматному преобразователю находить результат действия этого преобразователя на этой последовательности алгоритмически. Заметим также, что по произвольной морфической последовательности можно алгоритмически находить те символы, которые входят в неё бесконечно много раз.

Прямое доказательство следствия 14 приведено в [4].

В частности, ещё одним способом мы получаем доказательство того, что теории $\mathrm{MT}\langle \mathbf{t} \rangle$ и $\mathrm{MT}\langle \mathbf{f} \rangle$ разрешимы.

Литература

[1] J.-P. Allouche and J. Shallit. The ubiquitous Prouhet–Thue–Morse sequence. In *Sequences and their applications, Proceedings of SETA'98*, pages 1–16. Springer-Verlag, 1999.

[2] J.-P. Allouche and J. Shallit. *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, 2003.

[3] J. R. Büchi. On a decision method in restricted second-order arithmetic. In *Proceedings of International Congress for Logic, Methodology and Philosophy of Science*, pages 1–11. Stanford University Press, 1962.

[4] O. Carton and W. Thomas. The monadic theory of morphic infinite words and generalizations. *Information and Computation*, 176:51–65, 2002.

[5] F. M. Dekking. Iteration of maps by an automaton. *Discrete Mathematics*, 126(1–3):81–86, 1994.

[6] M. Lothaire. *Combinatorics on Words*. Cambridge University Press, 2nd edition, 1997.

[7] R. McNaughton. Testing and generating infinite sequences by a finite automaton. *Information and Control*, 9:521–530, 1966.

[8] M. Morse and G. A. Hedlund. Symbolic dynamics. *American Journal of Mathematics*, 60(4):815–866, 1938.

[9] M. Morse and G. A. Hedlund. Symbolic dynamics ii: Sturmian trajectories. *American Journal of Mathematics*, 62:1–42, 1940.

[10] An. Muchnik, A. Semenov, and M. Ushakov. Almost periodic sequences. *Theoretical Computer Science*, 304(1–3):1–33, 2003.

[11] Yu. Pritykin and M. Raskin. Almost periodicity and finite automata. In *Electronic Proceedings of WIWAD (satellite to CSR 2007), Ekaterinburg, Russia*, 2007.

[12] A. L. Semenov. Decidability of monadic theories. In *Proceedings of MFCS'84*, volume 176 of *Lecture Notes in Computer Science*, pages 162–175. Springer-Verlag, 1984.

[13] W. Thomas. Languages, automata, and logic. In G. Rozenberg and A. Salomaa, editors, *Handbook of Formal Languages*, volume 3, pages 389–455. Springer-Verlag, 1997.

[14] Ан. А. Мучник, Ю. Л. Притыкин, А. Л. Семёнов. Последовательности, близкие к периодическим. Препринт arXiv:0903.5316, 2009.

[15] Ю. Л. Притыкин. Конечно-автоматные преобразования строго почти периодических последовательностей. *Математические заметки*, 80(5):751–756, 2006.

[16] Ю. Л. Притыкин. Почти периодичность, конечно-автоматные преобразования и вопросы эффективности. *Известия ВУЗ. Математика*, принято к публикации. См. препринт `http://arXiv.org/abs/cs.DM/0607009`.

[17] М. А. Раскин. Об оценке регулятора автоматного образа почти периодической последовательности. *Труды XXVIII Конференции молодых учёных*, сс. 181–185, мех.-мат. ф-т МГУ им. Ломоносова, 2006.

[18] А. Л. Семёнов. Логические теории одноместных функций на натуральном ряде. *Известия АН СССР, серия математическая*, 47(3):623–658, 1983.