Автоматные слова

Анна Фрид

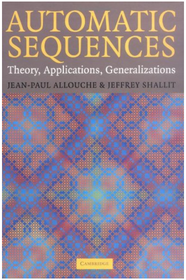
ИМ им. С. Л. Соболева СО РАН, Новосибирск
anna.e.frid@gmail.com

Лекция 2, 15.10.2011

Источник

Материал этой лекции в основном излагается по книге

Jean-Paul Allouche, Jeffrey Shallit, *Automatic Sequences — Theory, Applications, Generalizations*. Cambridge Univ. Press, 2003.



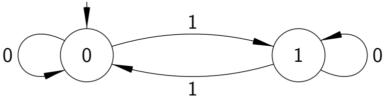
Материал из других источников в конце лекции будет отмечен отдельно.

Пример из жизни

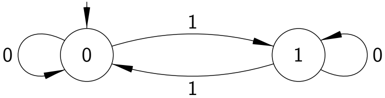
$$w = w_0 w_1 w_2 \cdots \in \{0,1\}^\omega,$$

$w_i =$ четность числа единиц в двоичном разложении числа i.

$$23 = [10111]_2 \Longrightarrow w_{23} = 1+1+1+1 \pmod 2 = 0.$$



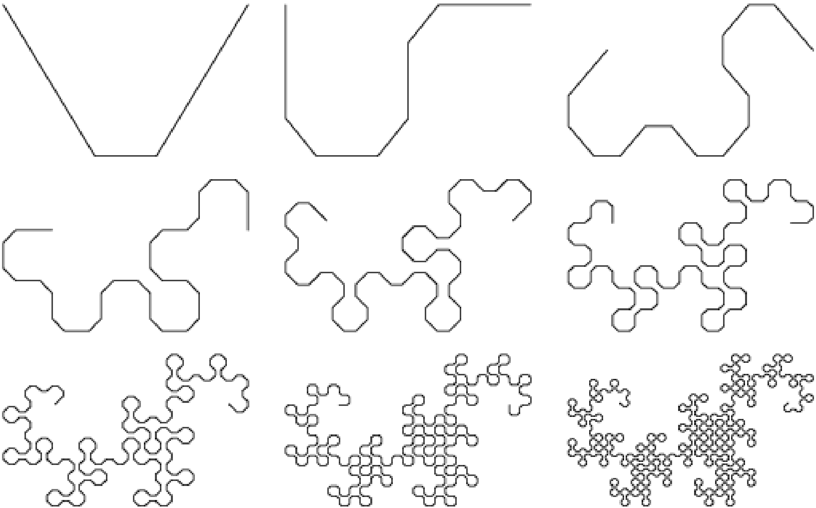
Пример из жизни



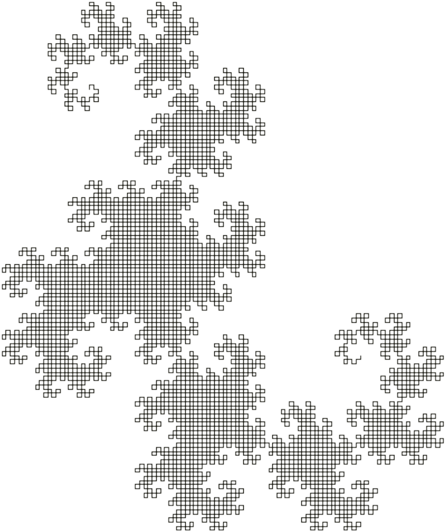
$$w = 0110100110010110 \cdots .$$

совершенно случайно это слово Туэ-Морса из предыдущей лекции

Еще один пример: сгибание бумаги



Кривая дракона



Слово дракона (paperfolding word)

$$w = w_1 w_2 w_3 \cdots \in \{1, -1\}^\omega$$

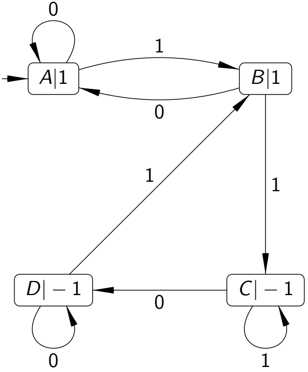
$$P = 1 \diamond -1 \diamond$$

$$\begin{array}{rcccccccccccccccc}
w(0) & = & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \\
w(1) & = & 1 & \diamond & \text{-1} & \diamond & 1 & \diamond & \text{-1} & \diamond & 1 & \diamond & \text{-1} & \diamond & 1 & \diamond & \text{-1} \\
w(2) & = & 1 & 1 & \text{-1} & \diamond & 1 & \text{-1} & \text{-1} & \diamond & 1 & 1 & \text{-1} & \diamond & 1 & \text{-1} & \text{-1} \\
w(3) & = & 1 & 1 & \text{-1} & 1 & 1 & \text{-1} & \text{-1} & \diamond & 1 & 1 & \text{-1} & \text{-1} & 1 & \text{-1} & \text{-1} \\
 & & & & & & & & & & & & & & & & \\
w & = & 1 & 1 & \text{-1} & 1 & 1 & \text{-1} & \text{-1} & 1 & 1 & 1 & \text{-1} & \text{-1} & 1 & \text{-1} & \text{-1}
\end{array}$$

Это код кривой дракона: 1 — налево, -1 — направо (или наоборот).

Автомат, выдающий слово дракона

$$w = w_{\color{red}1} w_2 w_3 \cdots = 1\;1\;\;-1\;1\;1\;\;-1\;\;-1\;1\;1\;1\;\;-1\;\;-1\;1\cdots$$



$$3 = [11]_2 \Longrightarrow w_3 = -1.$$

$$\Sigma_k = \{0, \ldots, k-1\}.$$

Definition

Слово $a = a_0 a_1 a_2 \cdots$ над алфавитом Δ называется k-автоматным для некоторого $k > 1$, если существует детерминированный конечный автомат с выходом (DFAO) $M = (Q, \Sigma_k, \delta, q_0, \Delta, \tau)$, такой что $a_n = \tau(\delta(q_0, w))$ для всех $n \geq 0$ и для всех слов $w \in \Sigma_k^*$, таких что $[w]_k = n$.

Другими словами, автомат получает на вход k-ичное разложение числа n и выдает символ a_n.

Lemma

Определение k-автоматного слова не зависит от того, как в автомат подается k-ичное разложение, слева направо или справа налево.

Доказательство. Возьмем DFAO $M = (Q, \Sigma_k, \delta, q_0, \Delta, \tau)$ из определения и определим новый автомат $M' = (S, \Sigma_k, \delta', q_0', \Delta, \tau')$ следующим образом:

- множество состояний S = множество всех функций g из Q в Δ;
- $q_0' = \tau$;
- $\tau'(g) = g(q_0)$ для всех $g \in S$;
- для всех $g \in S$, $a \in \Sigma_k$ определяем $\delta'(g, a) = h$, где функция h определяется равенствами $h(q) = g(\delta(q, a))$.

Далее индукция по длине входа.

Свойства автоматных слов

Lemma

Нули в начале k-ичного разложения тоже не мешают.

(Тоже требует доказательства!)

Свойства автоматных слов

k-волокно $I_k(w,d)$ слова $w = w_0 w_1 \cdots$: множество k-ичных разложений чисел n, таких что $w_n = d$.

$$I_k(w,d) = \{(n)_k : w_n = d\}.$$

Lemma

Слово w является k-автоматным тогда и только тогда, когда все его k-волокна — регулярные языки.

Свойства автоматных слов

Lemma

Слово, отличающееся от k-автоматного слова в конечном количестве позиций, k-автоматно.

Доказательство: через k-волокна, которые регулярны.

Свойства автоматных слов

t-периодическое слово: $w = uv^\omega$, где $|v| = t$.
Пример: слово $abbccc\ abac\ abac\ abac \cdots = abbccc(abac)^\omega$
4-периодично с предпериодом.

Lemma

Если слово t-периодическое (возможно, с предпериодом), то оно k-автоматно для любого k.

Доказательство: по предыдущей лемме, достаточно рассмотреть чисто периодический случай. Для него

$$Q = \{0, 1, \dots, t-1\};$$

$$\delta(q, b) = (kq + b) \pmod{t};$$

$$\tau(q) = w_q.$$

Свойства автоматных слов

Кодирование: отображение $\rho : \Delta \to \Delta'$.

Lemma

Слово, получающееся из k-автоматного применением кодирования, k-автоматно.

Доказательство: применяем кодирование к функции вывода.

Декартово произведение слов $w = w_0 w_1 \cdots$ и $v = v_0 v_1 \cdots$: слово $[w_0 \times v_0][w_1 \times v_1] \cdots$.

Lemma

Декартово произведение k-автоматных слов k-автоматно.

Доказательство: декартово произведение автоматов.

$$\boxed{q|a} \xrightarrow{\quad i \quad} \boxed{q'|a'}$$

$$\boxed{p|b} \xrightarrow{\quad i \quad} \boxed{p'|b'}$$

$$\boxed{[q \times p]|[a \times b]} \xrightarrow{\quad i \quad} \boxed{[q' \times p']|[a' \times b']}$$

Проверка на автоматность

Как доказать, что слово автоматно? Построить автомат.

Как доказать, что слово **НЕ** автоматно?
Стандартные методы:

- Через нерегулярность языка-волокна (pumping lemma etc.);
- Найти непериодическую последовательность вида $\{w_{[uv^i]_k}\}_{i=1}^{\infty}$;
- Другие методы — например, через частоты символов. В автоматном слове они, если существуют, должны быть рациональны (см. далее).

Морфизмы

Пусть Γ, Δ — конечные алфавиты. *Морфизмом* называется отображение $\varphi : \Gamma^* \to \Delta^*$, для всех слов $x, y \in \Gamma^*$ удовлетворяющее равенству $\varphi(xy) = \varphi(x)\varphi(y)$.

Ясно, что всякий морфизм однозначно определяется образами символов $\varphi(g)$, $g \in \Gamma$.

Example

Морфизм Фибоначчи $\varphi : \{0,1\}^* \to \{0,1\}^*$ $\varphi(0) = 01, \varphi(1) = 0$
Тогда, например, $\varphi(1101) = 00010$.

Однородные морфизмы и кодирования

Морфизм называется *(k-)однородным*, если длины образов всех символов алфавита Γ одинаковы и равны k.

1-однородный морфизм называется *кодированием*.

Example

$\varphi : a \mapsto ab, b \mapsto bc, c \mapsto ca$ — 2-однородный морфизм $\{a, b, c\}^* \to \{a, b, c\}^*$.

$\psi : a, c \mapsto 0, b \mapsto 1$ — кодирование.

Неподвижные точки морфизмов

$\varphi(a)$ начинается с a, длины слов $\varphi^k(a)$ неограниченно возрастают $\Longrightarrow$ у них существует предел, называемый *неподвижной точкой морфизма.*

$$a \to \varphi(a) \to \varphi^2(a) \to \varphi^3(a) \to \cdots \to \varphi^\infty(a)$$

Example

$$\varphi(0) = 01, \varphi(1) = 0$$

$$0 \to 01 \to 01\ 0 \to 010\ 01 \to 01001\ 010 \to 01001010\ 01001 \to \cdots$$

Предел: слово Фибоначчи

$$\varphi^\infty(a) = 0100101001001010010100100101001001 \cdots$$

Почему неподвижная точка? Потому что $\varphi(\varphi^\infty(a)) = \varphi^\infty(a)$.

CD0L-слова

CD0L-словом называется результат применения кодирования к неподвижной точке морфизма.

Example

$$\varphi : a \mapsto ab, b \mapsto bc, c \mapsto ca$$

$$\varphi^\infty(a) = abbcbccabccacaabbc\cdots$$

$$\psi : a, c \mapsto 0, b \mapsto 1$$

$$\psi(\varphi^\infty(a)) = 011010001000000110\cdots$$

Теорема Кобхэма

Theorem (Cobham,1972)

Бесконечное слово k-автоматно тогда и только тогда, когда является образом под действием кодирования неподвижной точки k-однородного морфизма.

Доказательство. Морфизмы строятся по автомату или наоборот. Пусть $w = \psi(\varphi^\infty(a))$, где $\varphi : \Delta \to \Delta^k$. Тогда

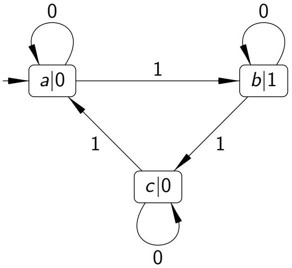
- множество состояний автомата равно Δ;
- начальное состояние - начальный символ a неподвижной точки морфизма;
- каждое состояние $b \in \Delta$ выдает символ $\psi(b)$;
- если $\varphi(b) = b_0 \cdots b_{k-1}$, то $\delta(b, i) = b_i$ для всех i.

$$\varphi : a \mapsto ab, b \mapsto bc, c \mapsto ca$$

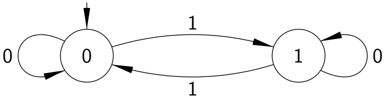
$$\varphi^\infty(a) = abbcbccabccacaabbc \cdots$$

$$\psi : a, c \mapsto 0, b \mapsto 1$$

$$\psi(\varphi^\infty(a)) = 011010001000000110 \cdots$$



Слово Туэ-Морса



$w = 0110\ 1001\ 1001\ 0110\ 1001 \cdots = \varphi^{\infty}(0)$, где

$$\varphi : \begin{cases} 0 \mapsto 01, \\ 1 \mapsto 10. \end{cases}$$

Большие значения k

Слово Керянена из предыдущей лекции 85-автоматно.

$$\varphi(a) = abcacdcbcdcadcdbdabacabadbabcbdbcbacbcdcacbab \\ dabacadcbcdcacdbcbacbcdcacdcbdcdadbdcbca$$

Много примеров (удобно строить).

Частоты символов

Частоты символов в неподвижной точке k-однородного морфизма (если они существуют) рациональны.

Следовательно, в k-автоматном слове — тоже.

Если частоты символов в слове существуют, но не рациональны, то слово не может быть k-автоматным.

Theorem (Honkala, 1986)

Вопрос о том, периодично ли (с предпериодом) k-автоматное слово, разрешим.

Theorem (Cassaigne,1994)

Для любого паттерна существует алгоритм проверки, избегает ли его данное циркулярное *k-автоматное слово.*

Техника Алуша-Шарлье-Рамперсада-Шаллита

Theorem (Charlier, Rampersad, Shallit, 2011)

Если свойство k-автоматного слова w выражаемо в терминах кванторов, логических операций, целочисленных переменных, операций сложения, вычитания, индексирования в слове w, и сравнения чисел или элементов слова w, то это свойство разрешимо.

Технику фактически впервые применил Lehr [1993], затем Allouche, Rampersad, Shallit [2009].

Пример применения техники

Перекрытие над алфавитом Σ — это слово вида $axaxa$, где $a \in \Sigma$, $x \in \Sigma^*$.

Theorem (Allouche, Rampersad, Shallit, 2009)

Следующий вопрос разрешим: избегает ли перекрытий данное k-автоматное слово w?

Пример применения техники

Theorem (Allouche, Rampersad, Shallit, 2009)

Следующий вопрос разрешим: избегает ли перекрытий данное k-автоматное слово w?

Доказательство. По автомату M_1, строящему слово, строим недетерминированный автомат M_2, который

- получает на вход k-ичные разложения чисел I и T;
- выясняет, существует ли число $J|\ 0 \leq J \leq T$ и $w_{I+J} \neq w_{I+J+T}$.

Если таких чисел нет, то $w_I \cdots w_{I+2T}$ — перекрытие.
Детерминизуем автомат M_2, получаем M_3.
Меняем местами принимающие и не принимающие состояния в M_3, получаем M_4.
В w есть перекрытия $\iff$ в M_4 есть достижимые принимающие состояния. □

Об автомате M_2

Множество состояний автомата M_2: пятерки вида

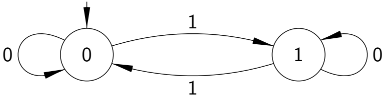
$$(b, c, d, q, r), \text{ где}$$

- $b \in \{<, =, >\}$ - отношение между J и T;
- c и d — carries при сложении очередных символов чисел $I + J$ и $I + T + J$ соответственно;
- q и r — состояния автомата M_1, соответствующие входам $I + J$ и $I + J + T$ соответственно.

Автомат недетерминированный: ребер каждого вида k штук, по количеству возможных (очередных) цифр числа J.

Для слова Туэ-Морса

В M_1 два состояния:



В M_2 состояний 3*2*3*2*2=72;
В детерминированном M_3 (и M_4) — 801 состояние.
Факт "слово Туэ-Морса избегает перекрытий" — самое сложное, что было реально проверенно чисто программным образом.

Другая идея

Можно строить автоматы, проверяющие свойства k-автоматных слов, не с помощью исходного автомата, а с помощью исходных морфизмов (по теореме Кобхэма).

Example

Рассмотрим k-автоматное слово $w = w_0 w_1 \cdots$ над алфавитом $\{0, 1, \ldots, d-1\}$ и обозначим через F_n k-ичную дробь

$$0, w_n w_{n+1} w_{n+2} \cdots .$$

Можно ли с помощью нового автомата определить, какое число больше, F_n или F_m?

Другая идея

Example

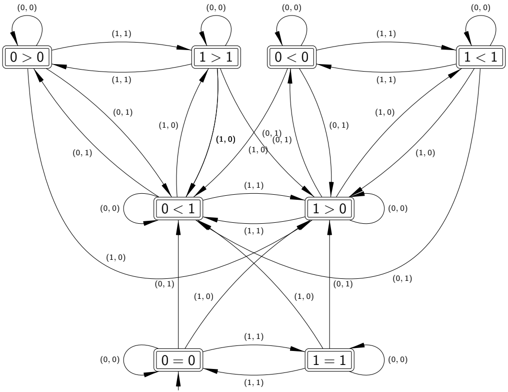
Рассмотрим k-автоматное слово $w = w_0 w_1 \cdots$ над алфавитом $\{0, 1, \dots, d-1\}$ и обозначим через F_n k-ичную дробь

$$0, w_n w_{n+1} w_{n+2} \cdots .$$

Можно ли с помощью нового автомата определить, какое число больше, F_n или F_m?

[Frid, Zamboni, 2010] Да;
у соответствующего автомата порядка $O((p^4)!)$ состояний, где p — мощность алфавита неподвижной точки k-однородного морфизма из теоремы Кобхэма.

У Туэ-Морса, как всегда, все хорошо:



Завтра

- Слова Штурма и дискретизация прямых. Свойства слов Штурма, эквивалентные определения. Быстрое построение слов Штурма.
- Количество всех слов Штурма: геометрический метод Берстеля и Поккьолы. Границы его применимости.
- Комбинаторные функции сложности бесконечных слов и факторных языков. Их свойства, методы их оценки.

Начало в 11.15