Math-Net.Ru

Общероссийский математический портал

Д. В. Пархоменко, Порожденные автоматами p -языки, *Дискрет. матем.*, 2014, том 26, выпуск 1, 96–102

DOI: <https://doi.org/10.4213/dm1270>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.234.156.22

26 апреля 2023 г., 06:11:45

Порожденные автоматами p -языки

© 2014 г. Д. В. Пархоменко¹

Ранее автор ввёл понятие p -языка как множества слов, возникающих на выходе некоторого автомата не менее p раз. Было показано, что каждый такой язык регулярен, но оставался вопрос, проверяемо ли свойство произвольного регулярного языка быть языком p -типа. Эта задача решена в данной статье.

Ключевые слова: автомат, раскраска автомата, регулярные языки, гистограммная автоматная функция.

Основные определения и результаты

Рассмотрим конечные алфавиты $A, B : A = \{a_1, \dots, a_{|A|}\}$, $B = \{b_1, \dots, b_{|B|}\}$ и конечный детерминированный инициальный автомат (см. [1]):

$$V = (A, Q, B, \phi, \psi, q_0). \quad (1)$$

Этот автомат порождает ограниченно детерминированную словарную функцию $f_V : A^* \rightarrow B^*$.

Точное определение и свойства автоматной функции описаны в [1]. Для простоты изложения в этой статье автор ограничился рассмотрением случая $|A| = |B| > 1$ (в случае, $|A| \neq |B|$ получаются аналогичные результаты).

Определение 1. Пусть задан конечный детерминированный инициальный автомат V . *Гистограммной автоматной функцией* автомата V назовем функцию $\kappa_V : B^* \rightarrow N \cup 0$, определенную формулой

$$\kappa_V(\beta) = |\{\alpha \in A^* | f_V(\alpha) = \beta\}|.$$

Функция κ_V каждому слову выходного алфавита сопоставляет мощность множества его прообразов при отображении f_V . Фактически, если использовать множество с кратными элементами $f_V(A^*)$, можно каждому слову из B^* поставить в соответствие его кратность в множестве $f_V(A^*)$. Другими словами, κ_V — это функция кратности выходных слов автомата V .

Определение 2. Произвольный непустой язык L алфавита B назовем *продолжаемым*, если для каждого слова $\beta \in B^*$ найдется такая буква $b \in B$, что $\beta b \in L$.

¹Место работы: LSI Corporation
e-mail: dcdenis@rambler.ru

Определение 3. Для произвольных конечного детерминированного автомата V и натурального p обозначим

$$L_p(V) = \{\beta \in B^* | \kappa_V(\beta) \geq p\}.$$

Таким образом, $L_p(V)$ — множество выходных слов, которые на выходе автомата V возникают не менее p раз. Регулярные языки с частотными свойствами ранее рассматривались в [2].

Рассмотрим вопрос о структуре множества выходных слов автомата. Полагая фиксированными алфавиты A, B , введём класс языков типа $\mathcal{L}_p$.

Определение 4. Для натурального p положим $\mathcal{L}_p = \{L_p(V) | V \in K(A, B)\}$, где $K(A, B)$ — класс всех автоматов вида $(A, Q, B, \phi, \psi, q_0)$.

В [3] установлено, что для любых конечного автомата V и натурального $p > 1$ язык $L_p(V)$ является регулярным и продолжаемым. Однако не каждое регулярное продолжаемое множество является языком типа $\mathcal{L}_p$.

При фиксированном натуральном p введем понятие p -раскраски автомата. Традиционно, рассмотрим события, представимые в автомате без выхода с помощью набора финальных состояний [4].

Определение 5. Пусть задан автомат

$$W = (B, Q, \phi, q_0, Q_F) \quad (2)$$

Введем функцию

$$k : Q \rightarrow N \cup \{0\},$$

сопоставляющую каждому состоянию из Q неотрицательное целое число. Будем говорить, что функция k задает на автомате V *правильную p -раскраску состояний*, если для любой вершины $q \in Q$ выполнены:

(1) $k(q) < p$ для любого $q \notin Q_F$, $k(q) \geq p$ для любого $q \in Q_F$ и $k(q_0) = 1$.

(2) $k(q) \cdot |B| = \sum_{b \in B} k(\phi(q, b))$, если $\phi(q, b) \notin Q_F$ для всех $b \in B$

$k(q) \cdot |B| \geq s \cdot p + \sum_{b \in B, \phi(q, b) \notin Q_F} k(\phi(q, b))$; где $s = |\{b \in B | \phi(q, b) \in Q_F\}|$, иначе.

Автомат V в таком случае назовём *правильно p -раскрашенным*. Ясно, что введенное понятие p -раскраски обобщается и на автоматы с бесконечным множеством состояний.

Заметим, что не всякий автомат типа (2) может быть правильно p -раскрашен.

Теорема 1. Пусть $p > 1$ — натуральное, а автомат $V' = (B, Q', \phi', q'_0, Q'_F)$ представляет некоторый язык L . Тогда $L \in \mathcal{L}_p$ в том и только том случае, когда найдётся правильно p -раскрашенный автомат $V = (B, Q, \phi, q_0, Q_F)$, также представляющий L , причем число состояний автомата V сравнимо с числом состояний исходного: $|Q| \leq \frac{p(p-1)}{2} \cdot |Q'| + p \cdot |Q'_F|$.

Следующий пример показывает, что увеличение числа состояний может оказаться необходимым. Для автоматов $V_1 = (E_3, Q, \phi_1, q_0, \{q_4\})$ и $V_2 = (E_3, S, \phi_2, s_0, \{s_4\})$ (соответственно, в верхней и нижней частях рис. 1) выполняется неравенство

$|Q| < |S|$. Автоматы V_1, V_2 порождают с помощью финальных состояний одинаковые регулярные языки из $\mathcal{L}_4$, однако V_2 обладает правильной 4-раскраской, а V_1 – нет. А именно, функция раскраски k автомата V_2 такова:

$$\begin{aligned} k(S_0) = 1, k(S_1) = 2, k(S_2) = 2, k(S_3) = 3, \\ k(S_4) = 4, k(S_5) = 0, k(S_6) = 1 \end{aligned}$$

Для V_1 построить 4-раскраску невозможно. Действительно, если бы правильная 4-раскраска k' для автомата V_1 существовала, то по определению $k'(q_0) = 1$, $k'(q_4) \geq 4$, следовательно, $k'(q_2) \geq 2$, так как $3k'(q_2) \geq k'(q_4) + 2k'(q_0) \geq 6$ и $k'(q_2) < 4$ ($q_2 \notin Q_F$). Аналогично из свойства 1) правильной раскраски следовало бы $3k'(q_1) = k'(q_0) + 2k'(q_2) = m$, где целое m равно либо 5, либо 7. Однако это равенство невозможно, так как функция k' правильной раскраски принимает только целочисленные значения.

Рис. 1. Автоматы V_1 и V_2 .

Следствие 1. Если язык L представим сильно связным автоматом, то $L \notin \mathcal{L}_p$ при всех $p > 1$.

Утверждение 1 доказывается с помощью мощностных соображений.

Утверждение 1. Если $V = (B, Q, \phi, q_0, Q_F)$ представляет некоторый язык $L \in \mathcal{L}_p$, то $p \leq |B||Q|^{-1}$.

Обозначим через $\text{Reg} = \text{Reg}(B)$ все регулярные языки над алфавитом B . Пусть $\mathcal{L}_0 = \text{Reg} \setminus \bigcup_{p=1}^{\infty} \mathcal{L}_p$. В [3] показано, что $\mathcal{L}_p \not\subseteq \mathcal{L}_q$ для любых натуральных $p \neq q$. Таким образом, система языков $\{\mathcal{L}_0, \mathcal{L}_1, \mathcal{L}_2, \dots\}$ образует избыточное покрытие множества регулярных языков в алфавите B .

Теорема 2. Существует алгоритм, который для любого языка $L \in \text{Reg}$ определяет все такие p , что $L \in \mathcal{L}_p$.

Доказательства

Формализуем определение источника, данное в [2], на языке множеств и отношений. Под источником понимается пятерка $V' = (B, Q, \lambda, q_0, Q_F)$, где B – входной

алфавит, Q — конечное множество состояний, отношение λ задано на декартовом произведении $Q \times Q \times B \times \{N \cup 0\}$, q_0 — начальное состояние, Q_F — множество финальных состояний. Если $(q_1, q_2, b, s) \in \lambda$, то в источнике имеется s кратных ребер (q_1, q_2) , которым приписан символ b . Таким образом, каждому пути в источнике единственным образом соответствует слово из B^* , однако одному слову может соответствовать несколько различных путей.

Множество слов, которым соответствуют пути, ведущие из q_0 в некоторое $q \in Q_F$, образуют язык, представляемый источником V' . Здесь и далее будем требовать, чтобы из каждого состояния источника выходило ровно $|B|$ ребер.

Пусть задан лес F , в котором каждому ребру каждого дерева приписана некоторая буква алфавита B .

Определение 6. Для любого натурального p назовем p -перечислимым множеством слов в источнике V' множество $L_p(V') = \{\beta \in B^* \mid \text{в } V' \text{ найдется не менее } p \text{ различных путей, которым соответствует слово } \beta \text{ и которые ведут из начального состояния } q_0 \text{ в одно из финальных состояний}\}$. Множество слов $L_p(F)$ назовем p -перечислимым лесом F , если $L_p(F) = \{\beta \in B^* \mid \text{в } F \text{ найдется не менее } p \text{ различных путей, с началом в вершинах деревьев леса, которым соответствует слово } \beta\}$.

Пусть заданы два леса F_1, F_2 и каждому ребру каждого дерева из этих лесов приписаны некоторые символы алфавита $B = \{b_1, \dots, b_{|B|}\}$.

Определение 7. Для произвольного натурального p назовем p -эквивалентными леса F_1 и F_2 , если $L_p(F_1) = L_p(F_2)$.

Легко проверить, что p -эквивалентные леса порождают одинаковые множества слов кратности большей или равной p .

Через D обозначим информационное дерево автомата V , ребрам которого приписаны только выходные символы алфавита B .

Доказательство Теоремы 1. Достаточность. Пусть автомат $V = (B, Q, \phi, q_0, Q_F)$ имеет правильную раскраску $k : Q \rightarrow Z_+$. Покажем, что существует такой автомат $V_1 = (B, Q_1, B, \phi_1, \psi_1, q_0')$, что $L = L_p(V_1)$. Очевидно, для этого достаточно доказать существование источника W с тем же свойством $L = L_p(W)$, у которого из каждого состояния выходит ровно $|B|$ ребер. Сопоставим каждому состоянию $q_i \in Q \setminus Q_F$ набор $\Omega(q_i) = \{q_i^1, \dots, q_i^{k(q_i)}\}$, а если $q_i \in Q_F$, то $\Omega(q_i) = \{q_i^1, \dots, q_i^p\}$. Элементы из введенных наборов будут состояниями источника W . Пусть, без ограничения общности, q_i таково, что:

$$\begin{aligned} \phi(q_i, b_1) &= \phi(q_i, b_2) = \dots = \phi(q_i, b_{s(1)}) = q_{t(1)}, \\ \phi(q_i, b_{s(1)+1}) &= \phi(q_i, b_{s(1)+2}) = \dots = \phi(q_i, b_{s(2)}) = q_{t(2)}, \\ &\dots \end{aligned}$$

$$\phi(q_i, b_{s(m-1)+1}) = \phi(q_i, b_{s(m-1)+2}) = \dots = \phi(q_i, b_{s(m)}) = q_{t(m)},$$

где $\sum_{i=1}^m s(i) = |B|$. Соединим $|B| \cdot k(q_i)$ ребрами состояния набора $\Omega(q_i)$ с состояниями наборов $\Omega(q_{t(1)}), \dots, \Omega(q_{t(m)})$, так, чтобы в каждое состояние набора $\Omega(q_{t(j)})$ для любого j входило одно ребро, если $q_{t(j)} \in Q \setminus Q_F$. В случае $q_{t(j)} \in Q_F$ потребуем, чтобы в первое состояние $q_{t(j)}^1$ набора $\Omega(q_{t(j)})$ входило не менее одного ребра. Также, если $q_i \in Q_F$, то первое состояние q_i^1 набора $\Omega(q_i)$ соединим только с первыми состояниями наборов, соответствующих финальным состояниям. Хотя бы одно такое состояние найдется, так как из условия существования правильной раскраски следует, что, если $q_i \in Q_F$, то существует $r : q_{t(r)} \in Q_F, r \leq m$. В противном случае из свойства 2 правильной раскраски следовало бы неравенство $k(q) \cdot |B| < p \cdot |B|$, а из

свойства 1 следовало бы равенство $k(q) \cdot |B| = p \cdot |B|$. Также будем требовать, чтобы из каждого состояния набора выходило ровно $|B|$ рёбер. Несложно видеть, что в случае существования правильной раскраски все эти требования выполнимы.

Итак, если состояния q_1, q_2 автомата V таковы, что $\phi(q_1, b) = q_2$, то в источнике W им соответствуют наборы $\Omega(q_1)$, и $\Omega(q_2)$, состоящие из $k(q_1)$ и $k(q_2)$ копий состояний q_1, q_2 соответственно. В каждое состояние набора $\Omega(q_2)$, за исключением, быть может, первого, входит единственное помеченное буквой b ребро, выходящее из состояний набора $\Omega(q_1)$. В первое же состояние набора $\Omega(q_2)$ входит одно ребро, если $q_2 \notin Q_F$, и не менее одного если $q_2 \in Q_F$. Значит, возможны 4 случая:

А) $q_1 \notin Q_F, q_2 \notin Q_F$. Пусть некоторое слово $\beta : \phi(q_0, \beta) = q_1$ ведет в каждое состояние набора $\Omega(q_1)$. В W не существует иных состояний, кроме принадлежащих этому набору, в которые вело бы слово β , так как таковых не существует и в автомате V . Принимая во внимание тот факт, что слово βb однозначно определяет состояние q_2 , а, следовательно, и набор $\Omega(q_2)$, получаем, что найдется ровно $|\Omega(q_2)| = k(q_2) < p$ путей в W , которым приписано слово βb .

В) $q_1 \in Q_F, q_2 \notin Q_F$. Пусть некоторое слово $\beta : \phi(q_0, \beta) = q_1$ таково, что в W существуют пути, которые соответствуют этому слову и ведут в каждое состояние набора $\Omega(q_1)$. В первое состояние q_1^1 может вести несколько таких путей. В W не существует иных состояний, в которые вело бы слово β , так как таковых не существует и в автомате V . По построению из первого состояния q_1^1 набора $\Omega(q_1)$ не выходит ребер, ведущих в какое-либо состояние набора $\Omega(q_2)$. Значит, в каждое состояние набора $\Omega(q_1)$, из которого выходят ребра в набор $\Omega(q_2)$, входит ровно один путь, соответствующий слову β . Следовательно, в W ровно $|\Omega(q_2)| = k(q_2) < p$ путей, которым приписано слово βb .

С) $q_1 \notin Q_F, q_2 \in Q_F$. Предположим, что в W существует ровно $k(q_1)$ различных путей, которым приписано слово β и которые ведут в каждое состояние набора $\Omega(q_1)$, где $\phi(q_0, \beta) = q_1$. Следовательно, в каждое состояние набора $\Omega(q_1)$ входит ровно один путь, соответствующий слову β . С другой стороны, из состояний набора $\Omega(q_1)$ выходит не менее $|\Omega(q_2)| = p$ ребер, по построению. Следовательно, число путей, соответствующих слову βb , не менее $|\Omega(q_2)| = p$.

Д) $q_1 \in Q_F, q_2 \in Q_F$. Абсолютно аналогично случаю В) получаем, что количество путей в W , соответствующих слову βb , не менее $|\Omega(q_2)| = p$.

Построение W таково, что если $q = \phi(q_0, b)$, то в каждое состояние набора $\Omega(q)$, кроме, быть может, первого, входит ровно одно ребро с пометкой b из $\Omega(q_0)$, а в первое состояние может входить несколько ребер. Значит $L_p(W) = L$. Заметим, что число состояний источника W , а следовательно, и автомата V' , не превосходит $p \cdot |Q|$.

Необходимость. Пусть теперь автомат $V' = (B, Q', \phi', q'_0, Q'_F)$ задает с помощью набора финальных состояний Q'_F язык $L \in \mathcal{L}_p$. Без ограничения общности, будем рассматривать автоматы без недостижимых состояний. Построим автомат V , представляющий язык L и имеющий правильную раскраску $k : Q \rightarrow \mathbb{Z}_+$. Из условия $L \in \mathcal{L}_p$ следует существование дерева $D : L = L_p(D)$. Каждому ребру дерева естественным образом приписана некоторая буква из B . Так как произвольное слово $\beta \in B^*$ однозначно определяет состояние $q' = \phi'(q'_0, \beta)$ автомата V' , то каждой вершине дерева D , в которую ведет путь из вершины, соответствующий слову β , можно однозначно сопоставить состояние q' автомата V' . Корню дерева D припишем q'_0 . Нагруженное дерево D задает частично определенную функцию $\hat{k} : B^* \times Q' \rightarrow \mathbb{N}$, где $\hat{k}(\beta, q')$ суть число вершин дерева, которым сопоставлено состояние q' и в которые ведут пути из корня, соответствующие слову β . Рассмотрим вершины дерева D , ко-

торым приписано состояние q' и в которые ведут пути по слову β из корня D . Пусть $F(\beta, q')$ — набор деревьев с корнями в этих вершинах. Тогда $\hat{k}(\beta, q')$ есть число корней деревьев, входящих в лес $F(\beta, q')$. Если лес $F(\beta, q')$ пуст, то считаем функцию $\hat{k}$ не определённой на наборе (β, q') . Далее, для заданного состояния q' допустимым словом будем называть такое слово $\beta \in B$, что $\phi'(q'_0, \beta) = q'$. Ясно, что хотя бы одно такое слово всегда найдется.

Для любого допустимого $\beta \in B$ выполнено: $0 < \hat{k}(\beta, q') < p$, если $q' \notin Q'_F$, и $\hat{k}(\beta, q') \geq p$, если $q' \in Q'_F$. Это верно, так как условие $q' \in Q'_F$ эквивалентно существованию в дереве D не менее p различных путей, которым приписано β .

Аналогично, для произвольных (β, q') сопоставим лесу $F(\beta, q')$ частично определённую функцию $k_{F(\beta, q')} : B^* \rightarrow N$, где $k_{F(\beta, q')}(\gamma)$ есть количество путей, исходящих из корней деревьев леса $F(\beta, q')$, которым приписано слово $\gamma \in B^*$. Заметим, что дерево D задает естественным образом частично определённую функцию $k : B^* \rightarrow N$, где $k(\delta)$ есть количество путей в D из его корня, которым приписано слово $\delta \in B^*$.

Заметим теперь, что равенство $\hat{k}(\alpha, q') = \hat{k}(\beta, q') < p$ для некоторого $q' = \phi(q_0, \alpha) = \phi(q_0, \beta)$ влечет за собой p -эквивалентность лесов $F(\alpha, q')$ и $F(\beta, q')$.

Аналогично покажем, что из $\hat{k}(\alpha, q') \geq \hat{k}(\beta, q') \geq p$ следует L_p -эквивалентность лесов $F(\alpha, q')$ и $F(\beta, q')$. Предположим, от противного, что найдется слово γ для которого $k_{F(\alpha, q')}(\gamma) \geq p$, но $k_{F(\beta, q')}(\gamma) < p$. Так как слова $\alpha, \beta \in L_p$, то $\phi'(q'_0, \alpha) = \phi'(q'_0, \beta) = q \in Q'_F$ по свойству автомата V' . Так как $k_{F(\alpha, q')}(\gamma) = k(\alpha\gamma) \geq p$, то $\phi'(q', \gamma) \in Q'_F$. Но, с другой стороны, $k_{F(\beta, q')}(\gamma) = k(\beta\gamma) < p$ и $\phi'(q', \gamma) \notin Q'_F$, что невозможно. Из полученного противоречия вытекает эквивалентность условий $k_{F(\alpha, q')}(\gamma) \geq p$ и $k_{F(\beta, q')}(\gamma) \geq p$ для произвольного γ .

Используя дерево D , построим автомат $V = (B, Q, \phi, q_0, Q_F)$, имеющий правильную p -раскраску и представляющий язык L . Состояниями автомата будут классы L_p -эквивалентности лесов дерева D . Для заданного состояния $q' \in Q' \setminus Q'_F$ в D существует не более $\frac{p(p-1)}{2}$ различных классов, а для $q' \in Q'_F$ — ровно один. Дерево D индуцирует соединение состояний нового автомата V , а именно, если $b \in B, \beta \in B^*$ таковы, что $\phi'(q'_0, \beta) = q, \phi(q, b) = q'$ и $\hat{k}(\beta, q) = k_1, \hat{k}(\beta b, q') = k_2$, то соединим ребром b состояния, заданные классами (q, k_1) и (q', k_2) . За начальное состояние q_0 возьмем класс, эквивалентный всему дереву D , т.е. класс $(q'_0, 1)$. Если пара (q', β) такова, что $\hat{k}(\beta, q') \geq p$, то класс лесов, эквивалентный $F(\beta, q')$, положим финальным состоянием нового автомата.

Для того, чтобы из каждого состояния выходило ровно $|B|$ ребер, дополним Q поглощающим состоянием, из которого переходы возможны только в себя. Введенное состояние не повлияет на язык, представляемый автоматом. Более того, условие $\beta \in B^*, \phi(q'_0, \beta) \in Q'_F$ эквивалентно условию $\hat{k}(\beta, q') \geq p$, что равносильно тому, что в автомате V путь, соответствующий слову β , ведет в состояние из Q_F . Получаем, что V представляет язык $L_p(D) = L$ посредством финальных состояний Q_F .

Пусть класс $\mathfrak{F}$ состоит из L_p -эквивалентных лесов дерева D . Выберем из $\mathfrak{F}$ любой лес $F(\beta', q')$, для которого значение $\hat{k}(\beta', q')$ минимально в классе. Обозначим это минимальное значение через m . Тогда для функции k : $k(\mathfrak{F}) = m$ автомат V обладает правильной p -раскраской. Теорема доказана.

Доказательство Следствия. Предположим обратное: найдутся такие $p > 1$ и автомат W , что $L = L_p(W)$. В силу сильной связности автомата W найдется такое $b \in B$, что значение гистограммной автоматной функции $\kappa_W(b) \geq p > 1$, в то время как для оставшихся букв алфавита B верно неравенство $\kappa_W(b') \geq 1$, где $b' \in B, b' \neq b$. Последнее неравенство выполнено, так как в силу сильной связности V

для любой буквы $b_1 \in B$ существует такое $b_2 \in B$, что $\kappa_W(b_1 b_2) \geq p$. Таким образом $\sum_{b \in B} \kappa_W(b) > |B|$. Но из мощностных соображений следует, что $\sum_{b \in B} \kappa_W(b) \leq |B|$. Противоречие.

Доказательство Теоремы 2. Рассмотрим некоторый регулярный язык L над алфавитом B . Пусть $V = (B, Q, Q_F, \phi, q_0)$ представляет L с помощью множества финальных состояний $Q_F \subset Q$. Пусть $n = |Q|$, тогда число различных автоматов вида (2) не превосходит $2^n \cdot n^{|B|}$. Согласно Теореме 1 проверка на принадлежность L к $\mathcal{L}_i$ требует исследования автоматов с не более чем $n \cdot i^2$ состояниями. То есть для проверки на принадлежность L к $\mathcal{L}_i$ требуется проверка не более чем $2^{n \cdot i^2} \cdot (n \cdot i^2)^{|B| \cdot n \cdot i^2}$ автоматов. Наконец, $L \notin \mathcal{L}_i$ для всех натуральных $i > |B|^{|Q|-1}$ согласно Утверждению 1. Теорема доказана.

Работа выполнена на кафедре МАТИС механико-математического факультета МГУ под руководством профессора, д.ф.-м.н. Д. Н. Бабина.

Список литературы

1. Кудрявцев В.Б., Алешин С.В., Подколзин А.С., *Введение в теорию автоматов*. Наука, Москва, 1985.
2. Бабин Д.Н., Холоденко А.Б., Об автоматной аппроксимации естественных языков. *Интеллектуальные системы* (2008) **12**, №3, 125–136.
3. Пархоменко Д.В., Спектральная автоматная функция и связанные с нею автоматные языки. *Интеллектуальные системы в производстве* (2012) **1**, №9, 165–175.
4. Клини С.К., Представление событий в нервных сетях и конечных автоматах. В сб.: *Автоматы*, Москва, 1956, с. 15–67.

Статья поступила 26.02.2013.