МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное автономное образовательное учреждение высшего образования

УРАЛЬСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ

имени первого Президента России Б. Н. Ельцина

ИНСТИТУТ ЕСТЕСТВЕННЫХ НАУК И МАТЕМАТИКИ

Кафедра алгебры и фундаментальной информатики

Различение слов перестановочными автоматами

Направление подготовки 02.04.02

«Фундаментальная информатика и информационные технологии»

Допустить к защите:

Зав. кафедрой:

доктор физико-математических наук

профессор М.В. Волков

_____

Нормоконтролер:

кандидат физимо-математических наук

доцент И.И. Иванов

_____

Магистерская диссертация

Карповой

Ольги Дмитриевны

_____

Научный руководитель:

доктор физико-математических наук,

профессор А.М. Шур

_____

Екатеринбург

2018

Содержание

ВВЕДЕНИЕ

Задача о различении двух наборов входных данных - одна из самых простых вычислительных задач, которые можно себе представить. Обычно, входные данные представлены в виде двух строк u и v над конечным алфавитом Σ и известны заранее.

В мощной вычислительной модели, такой как RAM, задача решается за константную память (помимо памяти, занимаемой строками): нам достаточно одного регистра для того, чтобы найти позицию, в которой различаются u и v.

Однако для более слабой модели, например, для конечных автоматов, задача существенно усложняется, и ее уже нельзя решить за константную память. Задача об определении минимального размера конечного автомата, различающего два данных слова, NP-трудна. Более того, не всё известно об асимптотике минимального количества вершин, необходимого, чтобы построить различающий автомат для каждой пары слов, длина которых меньше либо равна заданному наперед числу. Это и есть задача о различении слов автоматами, и на данный момент известны только верхняя и нижняя границы искомой функции, между которыми довольно большой разрыв.

Далее в работе будут даны все необходимые определения, задача будет поставлена более формально, также будет показана эквивалентность этой задачи поиску тождеств в полгруппе преобразований и группе перестановок. Более того, будут приведены серии тождеств, одна из которых позволяет слегка поднять нижнюю границу искомой функции.

1 Основные определения и постановка задачи

1.1 Основные определения

Пусть $\mathscr{A} = (\Sigma, Q, \delta, s, T)$ – детерминированный конечный автомат, где

Σ – входной алфавит, из которого формируются слова, принимаемые автоматом, $\Sigma \neq \emptyset$,

Q – множество состояний автомата, $Q \neq \emptyset$,

δ – функция переходов, определенная как отображение $\delta : Q \times \Sigma \to Q$,

s – начальное состояние, $s \in Q$

T – множество терминальных (конечных) состояний, $T \subseteq Q$.

Определение 1.1. Автомат принимает слово, если по окончании его обработки он находится в терминальном состоянии.

Определение 1.2. Пусть u и v – слова над алфавитом Σ. Говорят, что автомат различает слова u и v, если он принимает одно из них и не принимает другое.

Пример 1.1. Автомат на рисунке 1.1 различает слова 0010 и 1000.

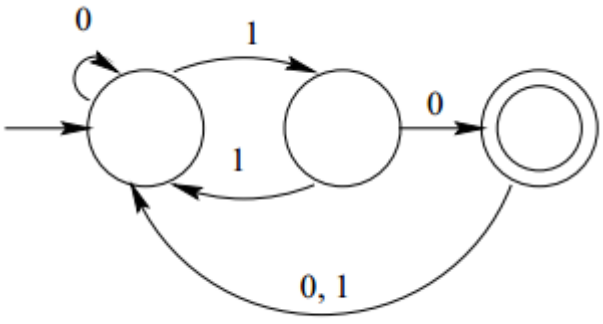

Рис. 1:

Определение 1.3. Пару слов (u, v) будем называть тождеством для некоторого автомата, если он либо принимает и u, и v, либо отвергает и u, и v, то есть не различает их.

Однако, в дальнейшем нам будет удобнее будет считать терминальными все состояния у рассматриваемых автоматов, и в связи с этим подкорректировать определения.

Определение 1.4. Автомат различает слова u и v, если он заканчивает их чтение в разных состояниях, то есть $s.u \neq s.v$.

Поскольку вся работа посвящена перестановочным автоматам, необходимо дать и это ключевое определение.

Определение 1.5. Автомат $\mathscr{A}$ – перестановочный, если переход из любого состояния по любому символу является перестановкой состояний, или, что тоже самое, для любого символа x из Σ и любых состояний q и p $\delta(q, x) \neq \delta(p, x)$.

1.2 Постановка задачи

Обозначим за $sep(u, v)$ количество состояний в минимальном детерминированном конечном автомате, различающем u и v, за $sep_p(u, v)$ – количество состояний в перестановочном автомате, различающем u и v.

Пример 1.2. Можно проверить, что ни один автомат из двух состояний не сможет различить слова 0010 и 1000, а как мы видели из рисунка 1.1, существует автомат с тремя состояниями, различающий эту пару. Значит, $sep(0010, 1000) = 3$.

Обозначим

$$S(n) = \max_{u \neq v; |u|,|v| \leq n} sep(u, v)$$

и

$$S_p(n) = \max_{u \neq v; |u|,|v| \leq n} sep_p(u, v)$$

Задача о различении слов автоматами, известная как Separating Words Problem, состоит в том, чтобы найти хорошую асимптотическую оценку

функций $S(n)$ и $S_p(n)$, то есть оценить, сколько состояний должно быть в автомате, различающем две строки длины n.

Эту задачу сформулировали Павел Горальчик и Вацлав Коубек в 1986 году [1, 2]. Они же доказали, что $S(n) = o(n)$. Позже этой задачей занимался Джон Робсон, который в 1989 году [3] доказал, что $S(n) = O(n^{2/5}(\log n)^{3/5})$ для произвольных автоматов, и в 1996 году [4] опубликовал статью, в которой было доказано, что $S_p(n) = O(n^{1/2})$.

1.3 Сведение к поиску тождеств в полугруппах и группах

Обозначим за T_k полугруппу всех отображений множества $1, ..., k$ в самого себя относительно операции композиции отображений. T_k называют полугруппой преобразований на k элементах.

Определение 1.6. Тождеством в полугруппе T называют пару слов (u, v) такую, что образы u и v под действием любого отображения $\Sigma \to T$ совпадают как элементы T. Длина тождества (u, v) - максимум длин u и v. Факт тождественности u и v в T_k будем обозначать как $u \equiv_k v$.

Определение 1.7. Полугруппа преобразований детерминированного конечного автомата $\mathscr{A}$ - это подполугруппа $T_{|Q|}$, состоящая из всех отображений $w : q \to q.w$, где $q \in \Sigma^*$.

Следующий факт соединяет между собой тождества в полугруппах и разделение слов автоматами.

Факт 1.1. *Для любой пары слов u, v, $u \equiv_k v \iff S(u, v) > k$*

Действительно, если $u \equiv_k v$, то это тождество выполняется и для полугруппы преобразований любого конечного детерминированного автомата с

k состояниями, откуда получаем $q.u = q.v$ для любого состояния q. С другой стороны, если для некоторого отображения $\rho : \Sigma \to T_k$ $\rho(u) \neq \rho(v)$ в T_k, тогда преобразования $\rho(a), a \in \Sigma$ могут быть использованы для создания автомата с k состояниями, отличающего u от v.

Известно, что задача о проверке $u \equiv_k v$ принадлежит классу coNP-complete для любого $k > 2$. Поэтому исходя из Факта 1.1, задача о проверке $S(u, v) \leq k$ является NP-полной. К тому же, задача о различении слов автоматами эквивалентна поиску асимптотики минимальной длины тождества в T_k.

До недавнего времени самым коротким тождеством в T_k было

$$x^{k-1} = x^{k-1+lcm(k)}, \tag{1}$$

где $lcm(k)$ - это наименьшее общее кратное всех чисел от 1 по k. Однако недавно было найдено новое тождество, которое короче 1, если k - простое или степень простого числа [5].

Имеет место аналогичная Факту 1.1 связь различения слов перестановочными автоматами и тождеств в группе перестановок S_k.

Факт 1.2. *Для любой пары слов u, v, $u \equiv_k v$ в S_k* $\iff$ $S_p(u, v) > k$

Самое короткое тождество в S_k, дающее почву для нижней оценки функции $S_p(k)$ имеет вид

$$x^{lcm(k)} = 1$$

Было несколько попыток найти более короткие групповые тождества. Существование тождества длины $O(e^{\sqrt{n \log n}})$ было доказано в статье [6]; основная идея заключается в применении функции Ландау о максимальном порядке перестановки. Совсем недавно было также доказано существование тождества длины $O(e^{\log^4 n \log \log n})$ [7], основанного на новых результатах о диаметре графа Кэли S_k. В данной работе будет представлена серия тождеств, показывающая, что $S_p(n) \geq \frac{3}{2} \log n + o(\log n)$.

1.4 Полезные факты

Вставить сюда ограничения: одинаковая длина слов, бинарный алфавит, начинаются и заканчиваются по-разному над алфавитом (xy) (yx)

2 Серии тождеств

Известно, что группа перестановок S_k удовлетворяет тождеству 1 и его бинарному собрату $x^{lcm(k)} \equiv_k y^{lcm(k)}$.

2.1 Тождетсва из двух блоков

2.2 Тождества из трех блоков

2.3 Тождества из четырех и более блоков

3 Заключение

Список литературы

[1] P. Goralčík and V. Koubek. On discerning words by automata // In L. Kott, editor, Proc. 13th Int'l Conf. on Automata, Languages, and Programming (ICALP), volume 226 of Lecture Notes in Computer Science, pages 116–122. Springer-Verlag, 1986.

[2] E.D. Demaine, S. Eisenstat, J. Shallit, D.A. Wilson. "Remarks on Separating Words // Descriptional Complexity of Formal Systems (DCFS 2011), P. 147-157. LNCS Vol. 6808. Springer, 2011.

[3] J. M. Robson. Separating strings with small automata // Inform. Process. Lett., 30:209–214, 1989.

[4] J.M. Robson. Separating words with machines and groups// RAIRO Inform. Théor. App. 30, 81–86 (1996)

[5] A.A. Bulatov, O. Karpova, A.M. Shur, K. Startsev. (2016). Lower Bounds on Words Separation: Are There Short Identities in Transformation Semigroups?. Electronic Journal of Combinatorics. 24.

[6] K. Bou-Rabee and D. B. McReynolds. Asymptotic growth and least common multiples in groups. Bull. Lond. Math. Soc., 43(6):1059-1068, 2011.

[7] G. Kozma and A. Thom. Divisibility and laws in

nite simple groups. Mathematische Annalen, 364(1):79-95, 2016.