О вычислении размаха регулярных языков при алфавитном кодировании

П. С. Дергач*

В данной работе вводится и изучается понятие размаха регулярных языков при алфавитном кодировании. Основным результатом является ранее не известное свойство дихотомии единичного и бесконечного размахов для языков специального итерационно-конкатанированного вида в случае двухэлементного входного алфавита. Также показано, как для такого алфавита на основании этого результата можно построить новый алгоритм вычисления размаха регулярного языка в общем случае (сама алгоритмическая разрешимость этого вычисления является известным фактом).

Ключевые слова: регулярный язык, алфавитное кодирование, размах.

1. Введение

В классической теории кодирования известен результат об алгоритмической разрешимости проверки регулярных языков на однозначность декодирования при алфавитном кодировании[1]. Также чуть менее известен результат об алгоритмичности вычисления максимального количества различных слов языка с общим кодом[2],[3]. Но соответствующие алгоритмы в общем случае имеют высокую временную сложность и имеет смысл изучать подклассы языков с точки зрения упрощения сложности алгоритмов и получения новых качественных результатов.

2. Основные определения

Через $\mathbb{N}$ обозначаем множество натуральных чисел. Понятия регулярного языка и алфавитного кодирования считаем общеизвестными [4], [5]. Множество регулярных языков в конечном алфавите A обозначаем через $R(A)$, а множество функций алфавитного кодирования из конечного

Дергач Петр Сергеевич — к.ф.-м.н., научный сотрудник кафедры Математической Теории Интеллектуальных Систем механико-математического факультета МГУ имени М. В. Ломоносова, e-mail: dergachpes@mail.ru.

Dergach Peter Sergeevich — Ph.D. in Physics and Mathematics, Researcher, Department of Mathematical Theory of Intelligent Systems, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

алфавита A в конечный алфавит B обозначаем через $F(A,B)$. *Размахом* $l_f(P)$ регулярного языка $P \in R(A)$ при алфавитном кодировании $f \in F(A,B)$ называем максимальное количество попарно различных слов из P с общим кодом относительно f. Если количество таких слов сколь угодно велико, то говорим, что *размах бесконечен* и пишем $l_f(P) = \infty$.

3. Основные результаты

Утверждение 1. *Пусть* $|A| = 2$, $s \in \mathbb{N} \cup \{0\}$, $P_1, \ldots, P_s \in R(A) \setminus \{\lambda\}$, $\alpha_1, \ldots, \alpha_{s+1} \in A^*$, $P = \alpha_1 \cdot P_1^* \ldots \alpha_s \cdot P_s^* \cdot \alpha_{s+1}$ *и* $f \in F(A,B)$. *Тогда или* $l_f(P) = 1$, *или* $l_f(P) = \infty$.

Замечание 1. *При* $|A| = 2$, $P \in R(A)$, $f \in F(A,B)$ *из этого утверждения следует алгоритмическая вычислимость* $l_f(P)$ *новым ранее не известным способом, который может в ряде случаев оказаться более быстрым, чем известные ранее алгоритмы.*

4. Краткое изложение идеи доказательства

Сначала заметим, что при $s = 0$ утверждение тривиально, а при $s = 1$ оно сводится к случаю $P = (P_1)^*$. В этом случае заметим, что если $l_f(P_1) > 1$, то найдется пара различных слов $\alpha, \beta \in P_1$ с общим кодом. Но тогда они не соизмеримы (то есть у них нет общего измельчения $\alpha = \nu^k$, $\beta = \nu^m$), а это в свою очередь значит, что при любом $n \in \mathbb{N}$ слова из множества $(\alpha \cup \beta)^n$ тоже имеют общий код и при этом их количество равно 2^n, то есть все такие комбинации из n слов α, β попарно различны. Далее считаем, что $s > 1$. Ясно, что $l_f(\alpha_1 \cdot P_1^* \ldots \alpha_s \cdot P_s^* \cdot \alpha_{s+1}) = l_f(P_1^* \ldots \alpha_s \cdot P_s^*)$. Далее используется ключевой факт о том, что при $A = \{0,1\}$ неоднозначное кодирование f без ограничения общности имеет вид $f(0) = 0^k$, $f(1) = 0^r$, где $k, r \in \mathbb{N}$. И для такого кодирования можно подобрать непустые слова $\alpha \in P_1^*$, $\gamma \in P_s^*$ так, чтобы $f(\alpha) = f(\gamma)$. И далее показывается, что если $\beta_1, \beta_2 \in P_1^* \ldots \alpha_s \cdot P_s^*$ – два различных слова с общим кодом, то каждая из двух пар слов $(\alpha\beta_1, \beta_1\gamma)$, $(\alpha\beta_2, \beta_2\gamma)$ имеет общий код и при этом хотя бы одна из этих пар попарно различна. Наконец, показывается, что если слова $\alpha\beta, \beta\gamma$ попарно различны и имеют общий код, то тогда и при любом фиксированном $n \in \mathbb{N}$ семейство слов вида $\alpha^{n-k}\beta\gamma^k$, $0 \leqslant k \leqslant n$ попарно различно и имеет общий код.

Приведем теперь обоснование первого замечания. Хорошо известно, что любой регулярный язык представим в виде конечного объединения множеств вида $\alpha_1 \cdot P_1^* \ldots \alpha_s \cdot P_s^* \cdot \alpha_{s+1}$. При этом у каждого представителя

этого объединения параметр s может быть своим и может равняться 0. И слова α_i тоже могут варьироваться и могут быть равными пустым словам. Регулярные множества P_i тоже могут варьироваться и являются непустыми. В силу первого утверждения каждая из таких серий имеет или единичный, или бесконечный размах. Причем можно эффективно быстро определить, какой из двух случаев имеет место, так как хорошо известен быстрый алгоритм проверки $l_f(P) \geqslant 2$. Если хотя бы одна из серий имеет бесконечный размах, то это верно и для P. В противном случае размах $l_f(P)$ конечен и не превосходит общего количества серий. Далее можно применить, например, модификацию графового критерия Маркова[1], в котором ищется цикл через пустую вершину. Основная суть модификации состоит в том, что в вершинах находятся уже не постфиксы элементарных кодов, а конструкции из n-ярусных слов, получаемых при подаче очередного элементарного кода из n-склейки (то есть набора из n различных слов с общим кодом). Также важно отметить, что помимо графового критерия здесь еще используется идея об отслеживании состояний автомата для P в каждом из ярусов и выявляется совпадение части возникающих слов между собой.

Заключение

Для случая $|A| = 2$ и языков интерационно-конкатанированного вида получен результат об эффективном вычислении размаха (все сводится к проверке обычной однозначности такого языка). Также показано, как этот результат может быть применен для построения алгоритма в случае $|A| = 2$ для произвольных регулярных языков. Дальнейший интерес представляет оптимизация сложности возникшего алгоритма и выявление классов языков, на которых данный алгоритм эффективно работает. Также было бы интересно получить обобщение этого алгоритма на случай $|A| > 2$.

Список литературы

[1] Ал. А. Марков, *Введение в теорию кодирования*, Наука, Москва, 1982, 192 с.

[2] Weber A., Seidl H., *On the degree of ambiguity of finite automata*, **88**, Theoretical Computer Science, 1991, 325-249 pp.

[3] Allauzen C., Mohri M., Rastogi A., *General algorithms for testing the ambiguity of finite automata*, Springer, Berlin, Heidelberg, 2008, 108–120 pp.

[4] В. Б. Кудрявцев, С. В. Алешин, А. С. Подколзин, *Теория автоматов: 2-е изд., испр. и доп.*, Юрайт, Москва, 2025, 320 с.

[5] С. В. Яблонский., *Введение в дискретную математику*, Высшая школа, Москва, 2008, 384 с.

Статья поступила 2 ноября 2026 г.

On the computation of the span of regular languages under alphabetic coding

P. S. Dergach

In this paper, we introduce and study the concept of the span of regular languages under alphabetic coding. The main result is a previously unknown dichotomy property of unit and infinite spans for languages of a special iteration-concatenation form in the case of a two-element input alphabet. It is also shown how, based on these results, one can construct an algorithm for computing the span in the general case for such an alphabet (the algorithmic decidability of this computation itself is a known fact).

Keywords: regular language, alphabetic coding, span.

References

[1] Markov Al.A., *Introduction to Coding Theory*, Nauka Publishers, Moscow, 1982, 192 pp.

[2] Weber A., Seidl H., *On the degree of ambiguity of finite automata*, **88**, Theoretical Computer Science, 1991, 325-249 pp.

[3] Allauzen C., Mohri M., Rastogi A., *General algorithms for testing the ambiguity of finite automata*, Springer, Berlin, Heidelberg, 2008, 108–120 pp.

[4] Kudryavtsev V.B., Alyoshin S.V., Podkolzin A.S., *Automata theory*, 2nd ed., revised and expanded, Urait, Moscow, 2025, 320 pp.

[5] Yablonskiy S.V., *Introduction to Discrete Mathematics*, Vysshaya Shkola, Moscow, 2008, 384 pp.

Received on November 2, 2026