Дискретная математика

том 6 выпуск 3 * 1994

УДК 519.12

Кодирование стохастических контекстно-свободных языков с однозначным выводом

© 1994 г. Л. П. Жильцова

Для стохастических контекстно-свободных языков с однозначным выводом исследуется задача оптимального кодирования, состоящая в минимизации математического ожидания длины закодированного слова в языке.

Рассмотрены класс кодирующих отображений, заданных на правилах порождающей грамматики, и класс конечно-автоматных отображений, для которых доказаны теоремы, аналогичные теореме Шеннона, о сколь угодно близком приближении в этих классах к оптимальному кодированию.

1. Введение

Объектом исследования в настоящей работе являются стохастические контекстно-свободные языки (стохастические КС-языки) с однозначным выводом, для которых рассматривается задача двоичного оптимального кодирования. Под оптимальным понимается кодирование на множестве слов языка, гарантирующее минимум математического ожидания длины закодированного слова.

Существование оптимального кодирования для стохастического КС-языка очевидно. Оптимальное кодирование состоит в кодировании слов языка в порядке убывания их вероятностей различными словами в двоичном алфавите, сначала словами длины 1, затем словами длины 2 и т. д. Однако такое кодирование и соответствующее ему декодирование не представляют практического интереса из-за сложности их реализации.

В работе рассмотрены два метода кодирования с полиномиальной временной сложностью реализации кодирования и декодирования, которые позволяют выбрать кодирование, сколь угодно близкое к оптимальному.

В первом методе осуществляется алфавитное кодирование выводов слов языка в порождающей грамматике. При этом правила грамматики, образующие слова, рассматриваются как буквы некоторого алфавита.

В §3 выводится конструктивное необходимое и достаточное условие существования стоимости алфавитного кодирования выводов слов стохастического КС-языка с однозначным выводом.

Стоимость алфавитного кодирования выводов существенно зависит от выбора грамматики, порождающей язык. В §2 приведен способ перехода от исходной грамматики к грамматике с «укрупненными» правилами, порождающей тот же самый язык. Последовательное «укрупнение» правил грамматики и применение обобщенно-префиксного кодирования [1] на множестве выводов слов дает асимптотически оптимальное кодирование для стохастического КС-языка с однозначным выводом (§4). Временная сложность такого кодирования сопоставима со сложностью распознавания КС-языков с однозначным выводом и не превосходит $O(n^2)$ операций, где n — длина кодируемого слова.

В §5 доказан аналогичный результат об асимптотически оптимальном кодировании для класса конечно-автоматных отображений на словах стохастического КС-языка с однозначным выводом. Алгоритм конечно-автоматного кодирования имеет, очевидно, линейную временную сложность.

Декодирование для обоих рассматриваемых методов кодирования имеет линейную временную сложность.

Заметим, что множество КС-языков с однозначным выводом содержит известные классы языков. Сюда относятся регулярные, линейные, детерминированные КС-языки. Для них рассматриваемые в работе классы кодирующих отображений имеют линейную временную сложность реализации кодирования.

Для регулярных языков оба класса рассматриваемых отображений совпадают, и полученные результаты об асимптотически оптимальном кодировании аналогичны теореме Шеннона о приближении в классе конечно-автоматных кодов к оптимальному кодированию [2]. Различие состоит в рассматриваемой вероятностной модели языка.

Результаты работы можно рассматривать как обобщение теоремы Шеннона на стохастические КС-языки с однозначным выводом.

2. Основные понятия и определения

Будем придерживаться определений контекстно-свободного языка (КС-языка) и стохастического КС-языка из [3, 4].

Стохастической порождающей КС-грамматикой будем называть систему $G = (V_T, V_N, R, S)$, где V_T и V_N — конечные множества терминальных и нетерминальных символов (терминалов и нетерминалов),

$$R = \bigcup_{i=1}^{n} R_i,$$

где R_i — конечное множество правил с одинаковой левой частью, имеющих вид

$$r_{ij}\colon A_i \xrightarrow{p_{ij}} \beta_{ij}, \qquad j = 1, \ldots, n_i,$$

где $A_i \in V_N$, $\beta_{ij} \in (V_N \cup V_T)^*$, а p_{ij} — вероятности применения правила r_{ij}, удовлетворяющие условиям $0 \leqslant p_{ij} \leqslant 1$ и $\sum_{j=1}^{n_i} p_{ij} = 1$.

Обозначим через $\Rightarrow$ отношение непосредственной выводимости, а через $\overset{*}{\Rightarrow}$ — рефлексивное транзитивное замыкание отношения $\Rightarrow$. Пусть $S \overset{*}{\Rightarrow} \alpha$. Для α будем рассматривать правый вывод в грамматике G, когда каждое правило в процессе вывода из S применяется к самому правому нетерминалу в слове. С правым выводом будем отождествлять соответствующую ему последовательность правил грамматики. В дальнейшем слово правый будем опускать, всегда подразумевая под выводом правый вывод слова в грамматике. Вывод слова α в грамматике G будем обозначать через $\omega(\alpha)$.

Пусть $\omega(\alpha) = r_{i_1j_1}r_{i_2j_2}\dots r_{i_kj_k}$ — вывод α, где $r_{i_lj_l}$ — правила грамматики. Вероятность вывода $p(\omega(\alpha))$ определим как произведение вероятностей входящих в $\omega(\alpha)$ правил:

$$p(\omega(\alpha)) = \prod_{i=1}^{k} p_{i_lj_l}.$$

Будем рассматривать грамматики с однозначным выводом [3].

Стохастический язык, порождаемый грамматикой G, есть

$$L(G) = \{(\alpha, p(\alpha)) \colon \alpha \in V_{T,}^{*},\ S \overset{*}{\Rightarrow} \alpha,\ p(\alpha) = p(\omega(\alpha))\}.$$

Таким образом, стохастический язык $L(G)$ определяется парой (L, P), где L — формальный язык, порожденный КС-грамматикой [3], а P — определенное на L распределение вероятностей. Формальный язык L будем называть характеристическим языком или просто языком.

Язык (L, P) называется согласованным, если $\sum_{\alpha\in L} p(\alpha) = 1$. Слову $\alpha \in L$ в грамматике с однозначным выводом соответствует единственное дерево вывода $D(\alpha)$ [3]. Высотой дерева вывода будем называть максимальную длину пути (т.е. число дуг) в дереве от корня до листа и обозначать ее через $d(\alpha)$. Назовем множество вершин, находящихся на расстоянии i от корня, i-м ярусом дерева. Дереву будем сопоставлять вероятность порождения слова в грамматике, соответствующего дереву.

Через L^i обозначим множество слов $\{\alpha \colon A_i \overset{*}{\Rightarrow} \alpha,\ \alpha \in V_T^*\}$.

Положим $L_t^i = \{\alpha \colon \alpha \in L^i,\ d(\alpha) \leqslant t\}$. Для слова α через $|\alpha|$ будем обозначать его длину.

Пусть $A_i \in V_N$. Через $I_1(A_i)$ обозначим множество нетерминальных символов, таких, что для любого $A_j \in I_1(A_i)$ существует слово $\alpha = \alpha_1 A_j \alpha_2 \in (V_N \cup V_T)^*$, выводимое из A_i, т.е. $A_i \overset{*}{\Rightarrow} \alpha$. Через $L_2(A_i)$ обозначим множество нетерминальных символов, таких, что для любого $A_j \in I_2(A_i)$ существует $\alpha = \alpha_1 A_i \alpha_2 \in (V_N \cup V_T)^*$, для которого $A_j \overset{*}{\Rightarrow} \alpha$. Таким образом, $I_1(A_i)$ — это множество нетерминалов, которые могут встретиться при выводе слов языка из A_i как аксиомы, а $I_2(A_i)$ — множество нетерминалов, при выводе слов из которых может встретиться символ A_i. Через $I_0(A_i)$ обозначим пересечение этих множеств, т.е. $I_0(A_i) = I_1(A_i) \cap I_2(A_i)$. Множество нетерминалов $K = \{A_{i_1}, \dots, A_{i_k}\}$, для которых $I_0(A_{i_j})$ совпадают и $I_0(A_{i_j}) \neq 0$, $j = 1, \dots, k$, назовем классом. Если $I_0(A_i) \neq 0$, будем считать, что A_i образует класс $\{A_i\}$. Для грамматики G множество V_N распадается на непересекающиеся классы. Грамматику G назовем неразложимой, если все нетерминалы из V_N образуют один класс. В противном случае G будем называть разложимой.

Введем в рассмотрение многомерные производящие функции $F_i(s_1, s_2, \dots, s_n)$, $i = 1, \dots, n$, где переменная s_i соответствует нетерминалу A_i. Функция $F_i(s_1, \dots, s_n)$ строится по множеству правил R_i с левой частью A_i следующим образом. Для каждого правила $A_i \xrightarrow{p_{ij}} \alpha_{ij}$ вводится слагаемое $q_{ij} = p_{ij}s_1^{k_1}s_2^{k_2}\dots s_n^{k_n}$, где k_l — число вхождений нетерминального символа A_l в правую часть правила, $l = 1, \dots, n$. Таким образом, $F_i(s_1, \dots, s_n) = \sum_{j=1}^{n_i} q_{ij}$.

Пусть

$$a_{ij} = \left.\frac{\partial F_i(s_1,\ldots,s_n)}{\partial s_j}\right|_{s_1=\ldots=s_n=1}$$

Квадратную матрицу A размера $n \times n$, образованную элементами a_{ij}, будем называть матрицей первых элементов. Обозначим через r максимальный по модулю собственный корень матрицы A. Матрица A неотрицательна, поэтому для нее существует максимальный действительный неотрицательный собственный корень (перронов корень) [5].

Приведем необходимое условие согласованности КС-языка с однозначным выводом, которое может быть получено интерпретацией результатов теории ветвящихся процессов применительно к процессу порождения слов в языке [6].

Для того, чтобы стохастический КС-язык $L(G)$ с однозначным выводом был согласованным, необходимо, чтобы перронов корень r матрицы A был меньше или равен 1.

В дальнейшем будем рассматривать согласованные стохастические КС-языки с однозначным выводом. Кроме того, будем считать, что грамматика G не содержит недостижимых нетерминалов, т.е. все нетерминалы используются при выводе слов языка [3].

Заметим, что регулярные и линейные языки всегда являются согласованными и для них $r < 1$ при отсутствии недостижимых нетерминалов.

В заключение параграфа опишем способ перехода от исходной грамматики G к эквивалентной грамматике G_t, используемый в дальнейшем при доказательстве основных результатов.

Пусть M_t^i — множество слов в алфавите $\{V_N \cup V_T\}$, выводимых из A_i, для которых высота дерева вывода не превосходит t, и нетерминалами могут быть помечены листья только t-го яруса дерева. Обозначим через L_t^i множество слов $\{\alpha\colon \alpha \in V_T^*,\ d(\alpha) \leqslant t\}$.

Теорема 1. *Справедливо равенство*

$$\sum_{\alpha \in M_t^i} p(\alpha) = 1.$$

Доказательство. Очевидно,

$$\sum_{\alpha \in M_t^i} p(\alpha) = \sum_{\alpha \in L_{t-1}^i} p(\alpha) + \sum_{\alpha \in M_t^i \setminus L_{t-1}^i} p(\alpha).$$

Первую и вторую сумму в правой части обозначим через Σ_1 и Σ_2 соответственно. Для слагаемых, входящих в Σ_2, представим $p(\alpha)$ в виде произведения $p(\alpha) = p(\beta)p(\alpha_1)\ldots p(\alpha_k)$, где β — слово, которому соответствует дерево вывода $D(\beta)$, получающееся из $D(\alpha)$ удалением вершин t-го яруса и инцидентных им дуг, а $\alpha_1,\ldots,\alpha_k$ — слова, соответствующие поддеревьям высоты 1 с корнями на t-м ярусе дерева $D(\alpha)$. Будем считать, что $\alpha_1,\ldots,\alpha_k$ упорядочены слева направо в соответствии с расположением корней их деревьев в $D(\alpha)$. Тогда

$$\Sigma_2 = \sum_{\alpha \in M_t^i \setminus L_{t-1}^i} p(\alpha) = \sum_{\beta,\alpha_1,\ldots,\alpha_k} p(\beta)p(\alpha_1)\ldots p(\alpha_k).$$

Сгруппируем члены суммы, имеющие одинаковые слова $\beta, \alpha_1, \ldots, \alpha_{k-1}$, тогда

$$\Sigma_2 = \sum_{\beta, \alpha_1, \ldots, \alpha_{k-1}} p(\beta) p(\alpha_1) \ldots p(\alpha_{k-1}) \sum_{\alpha_k} p(\alpha).$$

Учитывая, что $\sum_{\alpha_k} p(\alpha_k) = 1$, получим равенства

$$\Sigma_2 = \sum_{\beta} p(\beta) \sum_{\alpha_1, \ldots, \alpha_{k-1}} p(\alpha_1) \ldots p(\alpha_{k-1}) = \ldots = \sum_{\beta \in M_{t-1}^i \setminus L_{t-1}^i} p(\beta).$$

В итоге

$$\Sigma_1 + \Sigma_2 = \sum_{\alpha \in L_{t-1}^i} p(\alpha) + \sum_{\beta \in M_{t-1}^i \setminus L_{t-1}^i} p(\beta) = \sum_{\alpha \in M_{t-1}^i} p(\alpha).$$

Продолжая этот процесс, получим, что

$$\sum_{\alpha \in M_t^i} p(\alpha) = \sum_{\alpha \in M_{t-1}^i} p(\alpha) = \ldots = \sum_{\alpha \in M_1^i} p(\alpha).$$

Множество M_1^i есть не что иное, как множество правых частей правил из R_i. Поэтому

$$\sum_{\alpha \in M_1^i} p(\alpha) = \sum_{j=1}^{n_i} p_{ij} = 1.$$

Теорема доказана.

Используя множества M_t^i, мы можем перейти от грамматики G к грамматике G_t с множеством правил $R(t) = \bigcup_{i=1}^{n} R_i(t)$, где $R_i(t) = \{A_i \xrightarrow{p'_{ij}} \alpha'_{ij} : \alpha'_{ij} \in M_t^i\}$. Каждому правилу в $R_i(t)$ приписывается вероятность, равная вероятности вывода слова α_{ij} из A_i в исходной грамматике G. В силу доказанной теоремы G_t — стохастическая грамматика. Нетрудно показать, что $L(G) = L(G_t)$ и G_t — грамматика с однозначным выводом. Для G_t матрица первых моментов совпадает с t-й степенью матрицы A для G, т.е. равна A^t [6].

Если нетерминалы множества V_N образуют один класс, существует такое $t \geqslant 1$, для которого матрица A^t положительна [6]. Поэтому, не уменьшая общности, в дальнейшем будем полагать, что матрица первых моментов для неразложимой грамматики G положительна.

3. Условие существования стоимости алфавитного кодирования выводов

Пусть L — КС-язык и f — отображение L в $\{0,1\}^*$, удовлетворяющее требованию взаимной однозначности:

$$\text{если} \quad \alpha, \beta \in L \quad \text{и} \quad \alpha \neq \beta, \quad \text{то} \quad f(\alpha) \neq f(\beta). \qquad (1)$$

Класс всех отображений f, для которых выполняется условие (1), обозначим через $\vartheta(L)$.

Под стоимостью кодирования $C(L,f)$ будем понимать математическое ожидание длины слова $f(\alpha)$ в L:

$$C(L,f)=\sum_{\alpha\in L}p(\alpha)|f(\alpha)|=\lim_{N\to\infty}\sum_{d(\alpha)\leqslant N}p(\alpha)|f(\alpha)|.$$

и будем говорить, что $C(L,f)$ существует, если существует конечный предел величины $\sum_{d(\alpha)\leqslant N}p(\alpha)|f(\alpha)|$ при $N\to\infty$.

Пусть $\alpha\in L$. Рассмотрим вывод $\omega(\alpha)$ в грамматике G:

$$\omega(\alpha)=r_{i_1j_1}r_{i_2j_2}\dots r_{i_kj_k},\quad r_{i_lj_l}\in R_{i_l},\quad l=1,\dots,k.$$

Так как G — грамматика с однозначным выводом, между α и $\omega(\alpha)$ существует взаимно однозначное соответствие.

В качестве отображения f будем рассматривать алфавитное кодирование слов $\omega(\alpha)$, задаваемое схемой $\tilde f$:

$$r_{ij}\to v_{ij},\quad v_{ij}\in\{0,1\}^*,\quad i=1,\dots,n,\quad j=1,\dots,n_i.$$

Положим $f(\alpha)=\tilde f(\omega(\alpha))=v_{i_1j_1}v_{i_2j_2}\dots v_{i_kj_k}$.

Слово v_{ij} будем называть элементарным кодом, его длину c_{ij} — стоимостью правила r_{ij} ($c_{ij}=|v_{ij}|$), а сумму $C(R_i,f)=\sum_{j=1}^{n_i}p_{ij}c_{ij}$ — стоимостью кодирования множества правил R_i. Величину $c(\alpha)=|f(\alpha)|$ назовем стоимостью кодирования слова α.

Класс всех отображений f, задаваемых схемами $\tilde f$ на множестве правил и удовлетворяющих (1), обозначим через $\vartheta(G)$. Очевидно, $\vartheta(G)\subseteq\vartheta(L)$.

Для исходного языка L будем полагать $L=L^1$, т.е. $S=A_1$.

Теорема 2. *Пусть перронов корень r матрицы первых моментов грамматики G меньше 1. Тогда для любого $f\in\vartheta(G)$ математические ожидания $C(L^i,f)$, $i=1,\dots,n$, существуют и удовлетворяют системе уравнений:*

$$C(L^i,f)=C(R_i,f)+\sum_{j=1}^{n}a_{ij}C(L^j,f),\qquad i=1,\dots,n. \tag{2}$$

Для доказательства теоремы предварительно докажем несколько лемм.

Лемма 1. *Для любого $f\in\vartheta(G)$ и $t>1$ справедлива система неравенств*

$$C(L_t^i,f)\leqslant C(R_i,f)+\sum_{j=1}^{n}a_{ij}C(L_{t-1}^j,f),\qquad i=1,\dots,n,$$

где

$$C(L_t^i,f)=\sum_{\alpha\in L_t^i}p(\alpha)c(\alpha).$$

Доказательство. Рассмотрим

$$C(L_t^i,f)=\sum_{\alpha\in L_t^i}p(\alpha)c(\alpha)=C(L_1^i,f)+\sum_{\alpha\in L_t^i,\ d(\alpha)>1}p(\alpha)c(\alpha).$$

Представим $p(\alpha)$ в виде $p(\alpha) = p_{ij}p'_\alpha$, где p_{ij} — вероятность первого правила в выводе $\omega(\alpha)$, а $p(\alpha)'$ — произведение вероятностей остальных правил в выводе; аналогично $c(\alpha) = c_{ij} + c'(\alpha)$, где $c'(\alpha)$ — суммарная стоимость всех правил в выводе, кроме первого. Тогда

$$\begin{aligned} C(L_t^i, f) &= C(L_1^i, f) + \sum_\alpha p_{ij}p'(\alpha)(c_{ij} + c'(\alpha)) \\ &= C(L_1^i, f) + \sum_\alpha p_{ij}p'(\alpha)c_{ij} + \sum_\alpha p_{ij}p'(\alpha)c'(\alpha). \end{aligned}$$

Здесь суммирование ведется по всем словам α, удовлетворяющим условиям: $\alpha \in L_t^i$ и $d(\alpha) > 1$.

Рассмотрим отдельно две последние суммы, обозначив их через Σ_1 и Σ_2 соответственно. Сгруппируем слагаемые в Σ_1, относящиеся к словам из L_t^i, для которых совпадает первое правило вывода, тогда

$$\Sigma_1 = \sum_{j+1}^{n_i} p_{ij}c_{ij} \sum_{\alpha \in L_t^{ij}} p'(\alpha).$$

В этом равенстве через L_t^{ij} обозначено множество слов из L_t^i, для которых первое правило в выводе совпадает с r_{ij}.

Пусть правило r_{ij} содержит в правой части k_1 символов A_1, k_2 символов A_2, …, k_n символов A_n. Тогда $p'(\alpha)$ можно разбить на $k_1 + k_2 + \ldots + k_n$ сомножителей, каждый из которых соответствует слову, выводимому из некоторого нетерминала, образовавшегося после применения правила r_{ij}. Таким образом,

$$p'(\alpha) = p(\beta_1^1)p(\beta_2^1)\ldots p(\beta_{k_1}^1)\ldots p(\beta_1^n)p(\beta_2^n)\ldots p(\beta_{k_n}^n).$$

Очевидно, $d(\beta_k^l) \leqslant t - 1$ для всех $l = 1, \ldots, n$, $k = 1, \ldots, k_l$.

Сгруппируем в Σ_1 слагаемые, для которых все β_k^l совпадают, кроме последнего слова $\beta_{k_n}^n$, тогда

$$\Sigma_1 = \sum_j p_{ij}c_{ij} \sum_{(\beta_1^1, \ldots, \beta_{k_n-1}^n)} p(\beta_1^1)\ldots p(\beta_{k_n-1}^n) \sum_{\beta_{k_n}^n} p(\beta_{k_n}^n),$$

где последняя сумма есть суммарная вероятность слов из L_{t-1}^n. Обозначим ее через $P(L_{t-1}^n)$.

Продолжая далее группировать члены в Σ_1, получим, что

$$\Sigma_1 = \sum_j p_{ij}c_{ij}P^{k_1}(L_{t-1}^1)P^{k_2}(L_{t-1}^2)\ldots P^{k_n}(L_{t-1}^n).$$

Заметим, что $k_1, k_2, \ldots, k_n$ зависят от j, и $P(L_{t-1}^l) \leqslant 1$ для любого l, так как язык L^l согласованный, что нетрудно доказать, а $L_{t-1}^l \subseteq L^l$.

Рассмотрим

$$\Sigma_2 = \sum_\alpha p_{ij}p'(\alpha)c'(\alpha).$$

Представим $p'(\alpha)$ в виде произведения $k_1 + k_2 + \ldots + k_n$ сомножителей, как и для Σ_1, а $c'(\alpha)$ — в виде суммы того же числа слагаемых: $c'(\alpha) = c(\beta_1^1) + c(\beta_2^1) + \ldots +$

$c(\beta^1_{k_1}) + \ldots + c(\beta^n_1) + \ldots + c(\beta^n_{k_n})$. Подставим выражения для $p'(\alpha)$ и $c'(\alpha)$ в Σ_2 и после преобразований получим, что

$$\begin{aligned}\Sigma_2 &= \sum p_{ij} \sum_{\beta^1_1,\ldots,\beta^n_{k_n}} \sum_{l,k} p(\beta^1_1)\ldots p(\beta^n_{k_n}) c(\beta^l_{k_n}) \\ &= \sum_j p_{ij} \sum_{i,k} p(\beta^l_k) c(\beta^l_k) \sum_{\alpha_1,\ldots,\alpha_m} p(\alpha_1)\ldots p(\alpha_m),\end{aligned}$$

где последнее суммирование ведется по всем наборам $(\alpha_1,\ldots,\alpha_m)$, полученным из наборов $(\beta^1_1,\ldots,\beta^n_{k_n})$ исключением β^l_k. Оценивая последнюю сумму так же, как и $p'(\alpha)$ в Σ_1, получим, что

$$\begin{aligned}\Sigma_2 &= \sum_j p_{ij} \sum_{l,k} p(\beta^l_k) c(\beta^l_k) P^{k_1}(L^1_{t-1})\ldots P^{k_{l-1}}(L^{l-1}_{t-1}) \\ &\quad \times P^{k_l-1}(L^l_{t-1}) P^{k_{l+1}}(L^{l+1}_{t-1})\ldots P^{k_n}(L^n_{t-1}) \\ &= \sum_j p_{ij} \sum_{l,k} C(L^l_{t-1}) P^{k_1}(L^1_{t-1})\ldots P^{k_l-1}(L^l_{t-1})\ldots P^{k_n}(L^n_{t-1}) \\ &= \sum_j p_{ij} \sum_{l=1}^{n} k_l C(L^l_{t-1}) P^{k_1}(L^1_{t-1})\ldots P^{k_l-1}(L^l_{t-1})\ldots P^{k_n}(L^n_{t-1}).\end{aligned}$$

Подставляя Σ_1 и Σ_2 в $C(L^i_t, f)$, получаем равенство

$$\begin{aligned}C(L^i_t, f) &= C(L^i_1, f) + \sum_j p_{ij} c_{ij} P^{k_1}(L^1_{t-1})\ldots P^{k_n}(L^n_{t-1}) \\ &\quad + \sum_j p_{ij} \sum_{l=1}^{n} k_l C(L^l_{t-1}) P^{k_1}(L^1_{t-1})\ldots P^{k_l-1}(L^l_{t-1})\ldots P^{k_n}(L^n_{t-1}). \quad (3)\end{aligned}$$

Учитывая, что $P(L^l_{t-1}) \leqslant 1$ при любых l и $t-1$, можем записать неравенство

$$C(L^i_t, f) \leqslant C(L^i_1) + \sum_j p_{ij} c_{ij} + \sum_j p_{ij} \sum_{l=1}^{n} k_l C(L^l_{t-1}).$$

Индекс j здесь соответствует правилам из R_i, не являющимся заключительными, т.е. содержащим в правой части нетерминальные символы. Поэтому

$$C(L^i_1, f) + \sum_j p_{ij} c_{ij} = \sum_{j=1}^{n_i} p_{ij} c_{ij} = C(R_i), \qquad C(L^i_t, f) \leqslant C(R_i) + \sum_{l=1}^{n} C(L^l_{t-1}) \sum_j p_{ij} k_l.$$

Так как

$$\sum_j p_{ij} k_l = \sum_{j=1}^{n_l} p_{ij} k_l = a_{ij},$$

получаем систему неравенств

$$C(L^i_t) \leqslant C(R_i) + \sum_{l=1}^{n} a_{il} C(L^l_{t-1}).$$

Лемма доказана.

Лемма 2. *Для любого $f \in \vartheta(G)$ и $t > 1$ справедлива система неравенств*

$$C(L_t^i, f) \leqslant \sum_{l=1}^{n} \sum_{k=0}^{t-1} a_{ii}^{(k)} C(R_l, f), \qquad i = 1, \ldots, n,$$

где $a_{ii}^{(k)}$, $i = 1, \ldots, n$, — элементы матрицы A^k.

Доказательство. Проведем доказательство по индукции. При $t = 2$ по лемме 1

$$C(L_2^i, f) \leqslant C(R_i, f) + \sum_{l=1}^{n} a_{il} C(L_1^l, f) \leqslant \sum_{l=1}^{n} (\delta_{il} + a_{il}) C(R_l, f),$$

так как $C(L_1^l, f) \leqslant C(R_l, f)$. Здесь δ_{il} — элемент единичной матрицы,

$$\delta_{il} = \begin{cases} 1 & \text{при } i = l, \\ 0 & \text{при } i \neq l. \end{cases}$$

Предположим, что утверждение леммы верно для $t - 1$. Тогда для t

$$\begin{aligned} C(L_t^i, f) &\leqslant C(R_i, f) + \sum_{l=1}^{n} a_{il} C(L_{t-1}^l, f) \\ &\leqslant C(R_i, f) + \sum_{l=1}^{n} a_{il} \sum_{m=1}^{n} \sum_{k=0}^{t-2} a_{lm}^{(k)} C(R_m, f) \\ &= C(R_i, f) + \sum_{m=1}^{n} \sum_{k=0}^{t-2} \sum_{l=1}^{n} a_{il} a_{lm}^{(k)} C(R_m, f) \\ &= C(R_i, f) + \sum_{m=1}^{n} \sum_{k=0}^{t-2} a_{im}^{(k+1)} C(R_m, f) = \sum_{m=1}^{n} \sum_{k=0}^{t-1} a_{il}^{k} C(R_m, f). \end{aligned}$$

Лемма доказана.

Лемма 3. *При $r < 1$ числовой ряд $\sum_{t=0}^{\infty} a_{ij}^{(t)}$ сходится при любых $i = 1, \ldots, n$, $j = 1, \ldots, n$.*

Доказательство. Пусть грамматика G неразложима. Тогда при $t \to \infty$

$$a_{ij}^{(t)} = \sum_{l=n}^{n} d_l^{ij} \lambda_l^t = d_{ij} r^t + O(r_1^t),$$

где r — перронов корень матрицы A, λ_l — собственные значения матрицы A и $r_1 < r$ [6]. Для положительного числового ряда $\sum_{t=0}^{\infty} a_{ij}^{(t+1)}$

$$\lim_{t \to \infty} \frac{d_{ij} r^{t+1} + O(r_1^{t+1})}{d_{ij} r^t + O(r_1^t)} = r.$$

По признаку Даламбера ряд $\sum_{t=0}^{\infty} a_{ij}^{(t)}$ сходится при $r < 1$.

Для разложимой грамматики G либо $a_{ij}^{(t)} = 0$ для любого t, либо, начиная с некоторого t, $a_{ij}^{(t)} = \varphi_{ij}(t)r^t + O(\psi_{ij}(t)r_1^t)$, где $\varphi_{ij}(t)$ и $\psi_{ij}(t)$ — полиномы от t степени не выше n с положительными старшими коэффициентами, и $r_1 < r$ [6]. Поэтому результат о сходимости ряда $\sum_{t=0}^{\infty} a_{ij}^{(t)}$ остается верным. Лемма доказана.

Доказательство теоремы 2. Пусть $b_{ij} = \sum_{t=0}^{\infty} a_{ij}^{(t)}$. По лемме 2

$$C(L_t^i, f) \leqslant \sum_{l=1}^{n} \sum_{k=0}^{t-1} a_{il}^{(k)} C(R_l, f).$$

Так как $\sum_{k=0}^{t-1} a_{ij}^{(k)} \leqslant b_{il}$, то $C(L_t^i, f) \leqslant \sum_{l=1}^{n} b_{il} C(R_l, f)$. В силу леммы 3 при $r < 1$ величина $C(L_t, f)$ ограничена сверху, и поскольку $C(L_t^i, f)$ монотонно возрастает с ростом t, она имеет конечный предел.

В уравнении (3) из леммы 1 перейдем к пределу при $t \to \infty$, учитывая, что $\lim_{t\to\infty} P(L_{t-1}^i) = 1$ для любого i. Получим равенство

$$C(L_t^i, f) = C(L_1^i, f) + \sum_j p_{ij} c_{ij} + \sum_j p_{ij} \sum_{l=1}^{n} k_l C(L^l, f).$$

Здесь $C(L_1^i, f) + \sum p_{il} c_{ij} = C(R_i, f)$, а

$$\sum_j p_{ij} \sum_{l=1}^{n} k_l C(L^l, f) = \sum_{l=1}^{n} \sum_j p_{ij} k_l C(L^l, f) = \sum_{l=1}^{n} a_{ij} C(L^l, f).$$

Заметим, что для заключительных правил $k_l = 0$ для любого l, поэтому $\sum_j p_{ij} k_l = \sum_{j=1}^{n_i} p_{ij} k_l = a_{il}$, где суммирование в левой сумме ведется только по незаключительными правилам.

Окончательно получаем систему линейных уравнений для $C(L^i, f)$:

$$C(L^i, f) = C(R_i, f) + \sum_{l=1}^{n} a_{il} C(L^l, f), \qquad i = 1, \ldots, n.$$

Теорема 2 доказана.

Теорема 3. *Пусть перронов корень r матрицы A равен* 1. *Тогда $C(L^1, f)$ не существует для любого $f \in \vartheta(G)$.*

Доказательство. Предположим противное. Пусть $C(L^1, f)$ существует, т.е. $C(L_t^1, f)$ ограничена сверху.

Предварительно покажем, что если $C(L^1, f)$ существует, то $C(L^i, f)$ существует для любого i. Так как мы рассматриваем грамматику G без недостижимых нетерминалов, для любого нетерминала A_i существует слово в L, в процессе вывода которого появляется A_i, поэтому существует слово $\gamma = \alpha_1 A_i \alpha_2$, выводимое из аксиомы грамматики, где $\alpha_1, \alpha_2 \in V_t^*$. Пусть вероятность его вывода равна $p(\gamma)$. Рассмотрим подмножество слов $L' = \{\alpha_1\beta\alpha_2 : \beta \in L^i,\ \alpha_1, \alpha_2 \in V_t^*\}$ из L. Для него

$$\begin{aligned}
C(L', f) &= \sum_{\alpha \in L'} p(\alpha)c(\alpha) = \sum_{\beta \in L^i} p(\gamma)p(\beta)(c(\gamma) + c(\beta)) \\
&= p(\gamma)c(\gamma) \sum_{\beta \in L^i} p(\beta) + p(\gamma) \sum_{\beta \in L^i} p(\beta)c(\beta) \\
&= p(\gamma)c(\gamma) + p(\gamma)C(L^i, f), \\
C(L', f) &\leqslant C(L^1, f),
\end{aligned}$$

поэтому $C(L^1, f) \geqslant p(\gamma)c(\gamma) + p(\gamma)C(L^i, f)$ и в случае неограниченности $C(L^i, f)$ значение $C(L^1, f)$ также неограничено.

Перейдем от исходной грамматики G к грамматике G_k, описанной в §2. Тогда матрица первых моментов для G_k будет совпадать с k-й степенью матрицы A, а перронов корень матрицы A^k будет равен r^k [6].

В качестве кодирования для G_k возьмем прежнее кодирование f, которое обобщим на грамматику G_k естественным образом. Если первой части правила r'_{ij} грамматики G_k соответствует вывод $\omega = r_{i_1 j_1} r_{i_2 j_2} \ldots r_{i_l j_l}$ в грамматике G, положим $v'_{ij} = v_{i_1 j_1} v_{i_2 j_2} \ldots v_{i_l j_l}$, где v'_{ij} — элементарный код правила r'_{ij}. Для грамматики G_k преобразуем i-е уравнение из (3) в неравенство, отбросив в правой части уравнения первую сумму. Получим неравенство

$$C(L^1, f) \geqslant C(L^1, f) + \sum_{l=1}^{n} \sum_{j} p_{1j} k_l C(L_{t-1}^l, f) P^{k_1 - 1}(L_{t-1}^1, f) P^{k_2}(L_{t-1}^2) \ldots P^{k_n}(L_{t-1}^n).$$

В этом неравенстве все обозначения относятся к G_k, а член $C(L_k^1, f)$ соответствует тому, что для G_k множество слов из L, имеющих длину вывода, равную 1, совпадает с множеством слов из L, имеющих в исходной грамматике G длину вывода, не превосходящую k.

Так как по предположению $C(L_t^i, f)$ ограничена сверху и имеет предел для любого i, перейдем в неравенстве к пределу при $t \to \infty$. Получим, что

$$C(L^1, f) \geqslant C(L_k^1, f) + \sum_{l=1}^{n} \sum_{j} p_{1j} k_l C(L^l, f).$$

Преобразуем неравенство к виду

$$C(L^1, f) \geqslant C(L_k^1, f) + \sum_{l=1}^{n} a_{1j}^{(k)} C(L^l, f). \tag{4}$$

Неравенство справедливо для любого k. Рассмотрим случай неразложимой грамматики G. Для нее $a_{1l}^{(k)} = d_{1l} r^k + O(r_1^k)$, причем $d_{1l} > 0$ для любого l. Перейдем в (4) к пределу при $k \to \infty$. Тогда

$$C(L^1, f) \geqslant C(L^1, f) + \sum_{l=1}^{n} d_{1j} C(L^l, f).$$

Поскольку $\sum_{l=1}^{n} d_{1j} C(L^l, f) \geqslant c_0 > 0$ для некоторого c_0, получаем неравенство

$$C(L^1, f) \geqslant C(L^1, f) + c_0.$$

Получили противоречие.

Нетрудно провести аналогичные рассуждения и для случая разложимой матрицы, учитывая представление коэффициентов $a_{1l}^{(k)}$ при $k \to \infty$ [6] и тот факт, что $a_{1l}^{(k)} > 0$ для некоторого l. Теорема доказана.

Из теорем 2 и 3 вытекает необходимое и достаточное условие существования стоимости кодирования в классе $\vartheta(G)$, которое сформулируем в виде следующей теоремы.

Теорема 4. *Пусть L — КС-язык, порожденный грамматикой G с однозначным выводом, и $f \in \vartheta(G)$. Тогда стоимость кодирования существует в том и только в том случае, когда перронов корень r матрицы первых моментов грамматики G меньше* 1.

Замечание. При доказательстве теорем 2–4 не использовалось свойство взаимной однозначности отображения f на L. Поэтому эти теоремы справедливы для любого отображения f, заданного на множестве выводов слов из L и обладающего свойством: для $\omega = r_{i_1 j_1} r_{i_2 j_2} \dots r_{i_k j_k}$

$$|f(\omega)| = |f(r_{i_1 j_1})| + |f(r_{i_2 j_2})| + \dots + |f(r_{i_k j_k})|.$$

В дальнейшем будем полагать $r > 1$.

4. Асимптотически оптимальное кодирование выводов

Пусть L — КС-язык. Оптимальным кодированием на L будем называть отображение f_0, для которого

$$C(L, f_0) = \inf_{f \in \vartheta(L)} C(L, f)$$

Для $f \in \vartheta(L)$ величину $Q(L, f) = C(L, f) - C(L, f)$ назовем избыточностью кодирования f на L. Как отмечалось во введении, существует тривиальный алгоритм оптимального кодирования, который однако является переборным. В основе алгоритма лежит упорядочение слов языка L по убыванию их вероятностей. Будем считать, что слова с одинаковой вероятностью упорядочены лексикографически. Возможность упорядочения слов следует из того, что для $\alpha \in L$, имеющего вероятность $p(\alpha)$, все слова языка, которые могут стоять впереди слова α, имеют длину вывода, не превосходящую величины $d = \log p(\alpha) / \log p_{\max}$, где $p_{\max}$ — максимальная вероятность правила грамматики. Следовательно, для нахождения места слова α достаточно построить все слова с длиной вывода, не превосходящей d. Далее слова из L кодируются по порядку словами из $\{0,1\}^*$, сначала словами длины 1, затем словами длины 2, и т.д.

Построенное кодирование будем обозначать через f_0.

Рассмотрим способ кодирования, описанный в §3. Без ограничения общности будем полагать, что аксиома A_1 не встречается в правых частях правил грамматики G.

Пусть $f \in \vartheta(G)$. Из (2) следует, что значение стоимости кодирования $C(L^t, f)$ определяется значениями $C(R_i, f)$, т.е. стоимостью кодирования на множествах правил R_i.

Выберем в качестве порождающей грамматики для L грамматику G_k.

Рассмотрим кодирование $f_k \in \vartheta(G_k)$, которое строится следующим образом. Для любого i первоначально правила из $R_i(k)$ кодируются префиксным кодом Шеннона [2], для которого $c_{ij} =]-\log p_{ij}[$, где $]x[$ — ближайшее к x целое, не меньшее x. Здесь и далее основание логарифма равно 2. Обозначим такое кодирование через f_k^{SH}.

Рассмотрим незаключительные правила из $R_1(k)$. Так как высота дерева вывода для них в исходной грамматике равна k, число применяемых правил при выводе правой части любого правила из $R_1(k)$ не меньшее k, поэтому в грамматике G_k каждому незаключительному правилу соответствует вероятность, не превосходящая $p_{\max}^k$. Оценим снизу вероятность заключительного правила из $R_1(k)$ с высотой дерева вывода, равной l. Пусть $n_0 = \max_i\{n_i\}$, т.е. n_0 — максимальное число правил среди множеств R_i в грамматике G. Без ограничения общности можно считать $n_0 > 1$. В случае $n_0 = 1$ мы имеем либо пустой язык, либо язык, состоящий из единственного слова, что не представляет интереса для кодирования. Число правил, соответствующих дереву вывода высоты l, не превосходит $n_0 + n_0^2 + \ldots + n_0^l < n_0^{l+1} = m$ и вероятность для такого дерева не меньше, чем $p_{\min}^m$, где $p_{\min}$ — минимальное значение вероятности правила в грамматике G. Поэтому при $p_{\min}^m \geqslant p_{\max}^k$ каждое заключительное правило из $R_1(k)$ с высотой дерева вывода, не превосходящей l, имеет вероятность, не меньшую вероятности любого незаключительного правила из $R_1(k)$.

Оценим значение l, для которого выполняется последнее неравенство. Из неравенства следует, что $n_0^{l+1}\log p_{\min} \geqslant k\log p_{\max}$, откуда, учитывая, что $\log p_{\min} < 0$, получаем оценку $(l+1)\log n_0 \leqslant \log k + \log c_0$, где $c_0 = \log p_{\max}/\log p_{\min}$. Окончательно $l \leqslant \log k/\log n_0 + \log c_0/\log n_0 - 1$. Положим

$$l = \left[\frac{\log k}{\log n_0} + \frac{\log c_0}{\log n_0}\right]. \tag{5}$$

Преобразуем код f_k^{SH} в код f_k. Для этого заключительные правила из $R_1(k)$ расположим в порядке убывания их вероятностей и будем кодировать их различными последовательными словами из $\{0,1\}^*$, сначала словами длины 1, затем словами длины 2, и т.д. Такое изменение кода f_k^{SH} будем проводить до тех пор, пока вероятность заключительных правил остается не меньше $p_{\max}^k$. Очевидно, все заключительные правила из $R_1(k)$ с высотой дерева вывода, не превосходящей l, будут перекодированы. На остальных множествах правил $R_i(k)$, $k = 2, \ldots, n$, отображение f_k совпадает с f_k^{SH}. При переходе от f_k^{SH} к f_k длина элементарного кода правила может только уменьшиться, и перекодирование всегда может быть проведено таким образом, что никакие два правила не будут закодированы одинаково.

Полученный код является обобщенно-префиксным [1] на множестве выводов слов из L в грамматике G_k.

Обозначим через L_l множество слов из L, для которых дерево вывода в исходной грамматике G по высоте не превосходит значения l, заданного равенством (5).

Оценим избыточность для f_k. Нетрудно видеть, что

$$\begin{aligned} Q(L, f_k) &= C(L, f_k) - C(L, f_0) = C(L_l, f_k) + C(L \setminus L_l, f_k) - C(L_l, f_0) - C(L \setminus L_l, f_0) \\ &\leqslant C(L_l, f_k) + C(L \setminus L_l, f_k) - C(L_l, f_0). \end{aligned}$$

Для $\alpha \in L_l |f_k(\alpha)| = |f_0(\alpha)|$ в силу построения кодов f_0 и f_k, поэтому $C(L_l, f_k) = C(L_l, f_0)$. Для $\alpha \in L \setminus L_l$ $|f_k(\alpha)| \leqslant |f_k^{\text{SH}}(\alpha)|$, поэтому

$$C(L \setminus L_l, f_k) \leqslant C(L \setminus L_l, f_k^{\text{SH}} \leqslant C(L \setminus L_l, f_1^{\text{SH}}) = C(L, f_1^{\text{SH}}) - C(L_l, f_1^{\text{SH}}).$$

При выводе последнего неравенства использовался тот факт, что для любых k и $\alpha \in L$ справедливо равенство $|f_k^{\text{SH}}(\alpha)| = |f_1^{\text{SH}}(\alpha)|$, так как вероятность любого правила грамматики G_k представима в виде произведения вероятностей правил исходной грамматики G, а для произведения вероятностей $p_1 \ldots p_t$

$$\left]\log \frac{1}{p_1 \cdots p_t}\right[\leqslant \sum_{i=1}^{t} \left]\log \frac{1}{p_i}\right[.$$

Таким образом,

$$Q(L, f_k) \leqslant C(L, f_k) - C(L_l, f_0) + C(L, f_1^{\text{SH}}) - C(L_l, f_1^{\text{SH}}) = C(L_l, f_1^{\text{SH}}) - C(L_l, f_1^{\text{SH}}).$$

Пусть $k \to \infty$, тогда в силу (5) $t \to \infty$ и по теореме 1 $C(L_l, f_1^{\text{SH}})$ стремится к $C(L, f_1^{\text{SH}})$. Поэтому $Q(L, f_k) \to 0$.

Заметим, что вместо f_k можно было бы выбрать любой обобщенно-префиксный код, для которого вектор длин элементарных кодов не меньше аналогичного вектора для кода f_k.

Оценим сложность реализации кодирования и декодирования для f_k. При заданном k реализация $f_k(\alpha)$ для $\alpha \in L$ включает построение правого вывода $\omega(\alpha)$ в грамматике G_k и построение $f_k(\alpha) = \tilde{f}(\omega(\alpha))$. Временная сложность построения $\omega(\alpha)$ определяется сложностью задачи распознавания для L, и для КС-языков с однозначным выводом имеет порядок $O(|\alpha|^2)$ [3]. Задача построения $f_k(\alpha)$ по $\omega(\alpha)$ имеет линейную временную сложность, так как число правил в $\omega(\alpha)$ линейно зависит от $|\alpha|$.

Таким образом, получаем, что задача построения $f_k(\alpha)$ для $\alpha \in L$, где L — КС-язык с однозначным выводом, имеет временную сложность $O(|\alpha|^2)$.

Декодирование, т.е. расшифровка α по $f_k(\alpha)$, состоит из построения по $f_k(\alpha)$ правого вывода $\omega(\alpha)$ в грамматике G_k и из восстановления α по $\omega(\alpha)$. Нетрудно показать, что множество правых выводов слов КС-языка является детерминированным КС-языком. Для недетерминированного КС-языка задача декодирования обобщенно-префиксного кода, каковым является f_k, имеет линейную временную сложность [7], поэтому для восстановления $\omega(\alpha)$ по $f_k(\alpha)$ достаточно $O(\alpha)$ операций. Для расшифровки α по $\omega(\alpha)$ также требуется $O(\alpha)$ операций.

Таким образом, задачи кодирования и декодирования для $f_k(\alpha)$ имеют полиномиальную временную сложность. Окончательный результат можно сформулировать в виде следующей теоремы.

Теорема 5. *Пусть L — КС-язык, порожденный грамматикой G с однозначным выводом, для которой перронов корень матрицы A меньше 1. Тогда существует последовательность кодирующих отображений $\{f_1, f_2, \ldots, f_k, \ldots : f_k \in \vartheta(G)\}$ с полиномиальной временной сложностью реализации кодирования и декодирования, для которых избыточность кодирования $Q(f_k, L)$ стремится к 0 при $k \to \infty$.*

5. Конечно-автоматное кодирование

Рассмотрим возможности конечно-автоматного кодирования КС-языков с однозначным выводом. Обозначим класс конечно-автоматных отображений для L через $\vartheta(L)$. Пусть $B = \{b_1, b_2, \ldots, b_m\}$ — алфавит языка L и k — натуральное число.

Построим $f_k \in \vartheta_A(L)$ следующим образом. Выделим в L множество слов U_k, длина которых меньше k, и разложим эти слова в порядке убывания их вероятностей. Слова, имеющие одинаковые вероятности, упорядочим лексикографически. Закодируем слова из U_k по порядку словами из $\{0,1\}^*$: сначала словами длины 1, затем словами длины 2, и т.д.

К словам из L, длина которых превышает k, применим алфавитное кодирование, заданное префиксным кодом, для которого $|f_k(b_i)| = \,]\log m[$, $i = 1, \ldots, m$. Докажем, что f_k удовлетворяет требованию взаимной однозначности на L. Так как внутри множеств U_k и $L \setminus U_k$ кодирование взаимно однозначно в силу построения f_k, достаточно показать, что слова из U_k кодируются более короткими словами, чем слова из $L \setminus U_k$.

Для $\alpha \in L \setminus U_k$ справедливо неравенство $|f_k(\alpha)| \geqslant k]\log m[$. Число слов в $\{0,1\}^*$, имеющих длину, меньшую $k]\log m[$, равно $N_1 = 2 + 2^2 + \ldots + 2^{k]\log m[-1} = 2^{k]\log m[} - 2 \geqslant m^k - 2$, а число различных слов в языке L, длина которых меньше k, не превосходит $N_2 = m + m^2 + \ldots + m^{k-1} = (m^k - m)/m_1 \leqslant m^k - m$. Сравнивая N_1 и N_2, получаем, что $N_1 \geqslant N_2$. В силу построения f_k и U_k справедливо неравенство $|f_k(\alpha)| < k]\log m[$ для $\alpha \in U_k$, что дает возможность по длине закодированных слов отличать слова из U_k от слов из $L \setminus U_k$.

Таким образом, мы показали, что $f_k \in \vartheta(L)$. Очевидно, что кодирование f_k может быть реализовано конечным автоматом, так как для этого достаточно помнить конечное множество слов U_k и их коды, а для слов большей длины осуществлять алфавитное префиксное кодирование.

Оценим избыточность кодирования f_k. Ясно. что

$$Q(L, f_k) = C(L, f_k) - C(L_l, f_0) \leqslant C(U_k, f_k) - C(U_k, f_0) + C(L \setminus U_k f_k).$$

Заметим, что $C(U_k, f) - C(U_k, f_0) \leqslant 0$, поэтому

$$Q(L, f_k) \leqslant C(L \setminus U_k, f_k) = \sum_{|\alpha| \geqslant k} p(\alpha)|f_k(\alpha)| = \,]\log m[\sum_{|\alpha| \geqslant k} p(\alpha)|\alpha|.$$

Обозначим через c_0 максимальное число терминалов в правиле исходной грамматике G. Тогда $|\alpha| \leqslant c_0|\omega(\alpha)|$, где $\omega(\alpha)$ — вывод слова α в грамматике G, и

$$Q(L, f_k) \leqslant c_0]\log m[\sum_{|\alpha| \geqslant k} p(\alpha)|p(\alpha)|\omega(\alpha)|.$$

Оценим снизу высоту дерева вывода для слов, по которым ведется суммирование. Так как $|\omega(\alpha)| \geqslant k/c_0$, высота дерева вывода $d(\alpha)$ не меньше l_0, где l_0 определяется из равенства $1 + n_0 + n_0^2 + \ldots + n_0^{l_0} = [k/c_0]$, и n_0 — максимальное число нетерминалов в правиле грамматики. Поэтому

$$Q(L, f_k) \leqslant c_0] \log m[\sum_{d(\alpha) \geqslant l} p(\alpha)|\omega(\alpha)|.$$

Рассмотрим отображение f на множестве выводов слов из L, которое каждому правилу грамматики ставит в соответствие единицу. Тогда $|f(\omega(\alpha))| = |\omega(\alpha)|$ и

$$Q(L, f_k) \leqslant c_0] \log m[\sum_{d(\alpha) \geqslant l_0} p(\alpha)|f(\omega(\alpha))|.$$

Учитывая замечание к теоремам 2–4, получаем, что в случае, когда перронов корень матрицы A меньше 1, $C(L, f)$ существует. Поскольку

$$\sum_{d(\alpha) \geqslant l_0} p(\alpha)|a(\omega(\alpha))| = C(L, f) - C(L_t, f)$$

для $t = l_0 - 1$, то $l_0 \to \infty$ и $Q(L, f) \to 0$ при $k \to \infty$. Таким образом, доказана следующая теорема.

Теорема 6. *Пусть L — КС-язык, порожденный грамматикой с однозначным выводом, для которой перронов корень $r < 1$. Тогда существует последовательность конечно-автоматных кодирующих отображений $\{f_1, \ldots, f_k, \ldots : f_k \in \vartheta(L)\}$, для которых избыточность кодирования на L стремится к 0 при $k \to \infty$.*

Нетрудно показать, что декодирование для f_k также может быть реализовано конечным автоматом.

Список литературы

1. Марков А. А., Смирнова Т. Г. Алгоритмические основания обобщенно-префиксного кодирования. *Докл. АН СССР* (1984) **274**, №4, 790–793.

2. Шеннон К. *Работы по теории информации и кибернетике.* ИЛ, Москва, 1963.

3. Ахо А., Ульман Дж. *Теория синтаксического анализа, перевода и компиляции,* т. 1. Мир, Москва, 1978.

4. Фу К. *Структурные методы в распознавании образов.* Мир, Москва, 1977.

5. Гантмахер Ф. Р. *Теория матриц.* Наука, Москва, 1967.

6. Севастьянов В. А. *Ветвящиеся процессы.* Наука, Москва, 1971.

7. Жильцова Л. П. Об алгоритмической сложности задач оптимального кодирования для контекстно-свободных языков. *Дискретная математика* (1989) **1**, №2, 38–51.

Статья поступила 03.10.92.