ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Июль—сентябрь 2003. Серия 1. Том 10, № 3, 23–53

УДК 519.713

ЗАКОНОМЕРНОСТИ В ДЕРЕВЬЯХ ВЫВОДА СЛОВ СТОХАСТИЧЕСКОГО КОНТЕКСТНО-СВОБОДНОГО ЯЗЫКА И НИЖНЯЯ ОЦЕНКА СТОИМОСТИ КОДИРОВАНИЯ. КРИТИЧЕСКИЙ СЛУЧАЙ*)

Л. П. Жильцова

Рассматривается язык, порожденный стохастической контекстно-свободной грамматикой, матрица первых моментов которой неразложима, непериодична и ее перронов корень равен 1. Для такого языка установлены закономерности в деревьях вывода фиксированной высоты t при $t \to \infty$. На основе этих закономерностей получена точная нижняя оценка стоимости двоичного кодирования.

Автором в [3, 4] рассматривались вопросы, связанные с кодированием сообщений, являющихся словами стохастического контекстно-свободного языка (стохастического КС-языка), при условии, что матрица первых моментов грамматики неразложима, непериодична и ее максимальный по модулю собственный корень (перронов корень) строго меньше единицы.

В настоящей статье рассматриваются аналогичные вопросы для случая, когда перронов корень неразложимой и непериодической матрицы первых моментов равен единице. По аналогии с теорией ветвящихся процессов этот случай будем называть критическим.

Для стохастического КС-языка в качестве слов большой длины рассматриваются слова, каждое из которых задано деревом вывода высоты t. При $t \to \infty$ найдено математическое ожидание числа применений произвольного правила грамматики на фиксированном ярусе дерева вывода, а также математическое ожидание числа применений правила для всего дерева вывода.

*)Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проект 01–01–00464).

На основе найденных закономерностей в применении правил грамматики получена точная нижняя оценка стоимости кодирования для КС-языка с однозначным выводом в критическом случае.

1. Основные определения

Стохастической КС-грамматикой называется система $G = \langle V_T, V_N, R, s\rangle$, где V_T и V_N — конечные множества терминальных и нетерминальных символов (терминалов и нетерминалов) соответственно; $s \in V_N$ — аксиома, $R = \bigcup_{i=1}^{k} R_i$, где k — мощность алфавита V_N и $R_i = \{r_{i1}, \dots, r_{i,n_i}\}$ — множество правил с одинаковой левой частью A_i. Каждое правило r_{ij} из R_i имеет вид

$$r_{ij} : A_i \xrightarrow{p_{ij}} \beta_{ij}, \quad j = 1, ..., n_i,$$

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

Для слов α и β из $(V_T \cup V_N)^*$ будем говорить, что β *непосредственно выводимо* из α (и записывать $\alpha \Rightarrow \beta$), если $\alpha = \alpha_1 A_i \alpha_2$, $\beta = \alpha_1 \beta_{ij} \alpha_2$ для некоторых $\alpha_1, \alpha_2 \in (V_T \cup V_N)^*$, и в грамматике G имеется правило $A_i \xrightarrow{p_{ij}} \beta_{ij}$.

Обозначим через $\Rightarrow_*$ рефлексивное транзитивное замыкание отношения $\Rightarrow$. Язык L_G, порождаемый грамматикой G, определяется как множество слов $\{\alpha |\, s \Rightarrow_* \alpha, \alpha \in V_T^*\}$.

Пусть $s \Rightarrow_* \alpha$. *Левым выводом* слова α назовем вывод, в котором каждое правило в процессе вывода слова α из аксиомы s применяется к самому левому нетерминалу в слове. Последовательность правил в левом выводе будем обозначать через $\omega(\alpha)$.

Важное значение имеет понятие дерева вывода [1]. Дерево строится следующим образом.

Корень дерева помечается аксиомой s. Пусть при выводе слова α на очередном шаге в процессе левого вывода применяется правило $A \xrightarrow{p_{ij}} b_{i_1} b_{i_2} \dots b_{i_m}$, где $b_{i_l} \in V_N \cup V_T$ $(1 \leqslant l \leqslant m)$. Тогда из самой левой вершины-листа дерева, помеченной символом A (при обходе листьев дерева слева направо), проводится m дуг в вершины следующего яруса, которые помечаются слева направо символами $b_{i_1}, b_{i_2}, \dots, b_{i_m}$ соответственно. После построения дуг и вершин для всех правил грамматики в выводе слова языка все листья дерева помечены терминальными символами и само слово получается при обходе листьев дерева слева направо.

Высотой дерева вывода будем называть максимальную длину пути от корня к листу.

Пример. Рассмотрим грамматику $G_0 = \langle \{x, \bar{x}\}, \{N\}, R, N\rangle$, в которой множество R состоит из двух правил:

$$r_1 : N \xrightarrow{p} xN\bar{x}N,$$
$$r_2 : N \xrightarrow{1-p} \lambda \ (\lambda\text{— пустое слово}).$$

Грамматика G_0 порождает известный язык Дика.

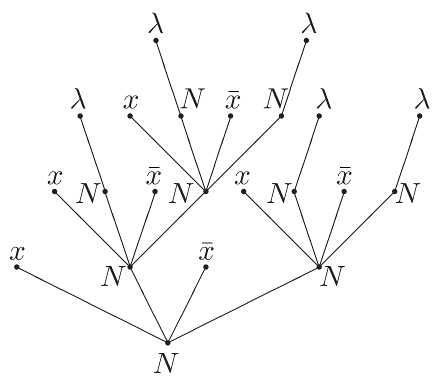

Рис. 1

На рис. 1 изображено дерево вывода в грамматике G_0. Ему соответствует левый вывод $r_1r_1r_2r_1r_2r_2r_1r_2r_2$ и слово $\alpha = xx\bar{x}x\bar{x}\bar{x}x\bar{x}$. Высота дерева вывода равна 4.

Пусть $\omega(\alpha) = r_{i_1j_1}r_{i_2j_2}\dots r_{i_nj_n}$ — некоторый левый вывод слова $\alpha \in L$ и d_α — соответствующее ему дерево вывода. Определим $p(d_\alpha)$ как произведение вероятностей правил, образующих $\omega(\alpha)$, т. е. $p(d_\alpha) = p_{i_1j_1}p_{i_2j_2}\dots p_{i_nj_n}$. Вероятность появления слова α определим как $p(\alpha) = \sum p(d_\alpha)$, где суммирование ведется по всем различным деревьям вывода слова α.

Грамматика G называется *согласованной*, если $\sum\limits_{\alpha\in L_G} p(\alpha) = 1$. В дальнейшем будем рассматривать согласованные КС-грамматики. Согласованная КС-грамматика G индуцирует распределение вероятностей P_G на множестве слов языка L_G. Язык L, порожденный согласованной стохастической КС-грамматикой, с распределением вероятностей P_G на L будем называть *стохастическим* КС-языком.

В дальнейшем важное значение будет иметь матрица первых моментов, которая определяется следующим образом. Рассмотрим многомерные производящие функции

$$F_i(s_1, s_2, \dots, s_k),\ 1 \leqslant i \leqslant k,$$

где переменная s_i соответствует нетерминальному символу A_i [5]. Функция $F_i(s_1, s_2, \dots, s_k)$ строится по множеству правил R_i с одинаковой левой частью A_i следующим образом. Для каждого правила $A_i \xrightarrow{p_{ij}} \beta_{ij}$ выписывается слагаемое

$$q_{ij=}p_{ij}s_1^{l_1}s_2^{l_2}\dots s_k^{l_k},$$

где l_m — число вхождений нетерминального символа A_m в правую часть правила ($1 \leqslant m \leqslant k$). Тогда $F_i(s_1, s_2, \dots, s_k) = \sum\limits_{j=1}^{n_i} q_{ij}$. Пусть

$$a_{ij} = \left.\frac{\partial F_i(s_1, \dots, s_k)}{\partial s_j}\right|_{s_1=s_2=\dots=s_k=1}.$$

Квадратная матрица A порядка k, образованная элементами a_{ij}, называется *матрицей первых моментов* грамматики G. Поскольку матрица A неотрицательна, существует максимальный по модулю действительный неотрицательный собственный корень (перронов корень) [2]. Этот корень обозначим через r.

В дальнейшем будем рассматривать грамматики, матрицы первых моментов которых неразложимы и непериодичны [2] и $r = 1$. Для неразложимой непериодической матрицы правый и левый собственные векторы, соответствующие перронову корню, могут быть выбраны положительными [2]. Через $U = (u_1, \dots, u_k)$ и $V = (v_1, \dots, v_k)$ будем обозначать положительные правый и левый собственные векторы, соответствующие перронову корню. Будем полагать, что выполняется нормировка $\sum\limits_{i=1}^{k} u_i v_i = 1$.

С помощью производящих функций определим вторые моменты. Вторым моментом будем называть величину

$$b_{ijm} = \left.\frac{\partial^2 F_i(s_1, \dots, s_k)}{\partial s_j \partial s_m}\right|_{s_1=s_2=\dots=s_k=1} \quad (i, j, m \in \{1, 2, \dots, k\}).$$

В дальнейшем будем полагать, что $b_{ijm} \neq 0$ при некоторых $i, j, m \in \{1, \dots, k\}$. Это означает, что в грамматике существуют правила, содержащие в правой части более одного нетерминала.

Пусть $\beta = (\beta_1, \dots, \beta_k)$ — неотрицательный целочисленный вектор и $p^i_\beta(t)$ — вероятность появления деревьев вывода с корнем, помеченным нетерминалом A_i, в каждом из которых на ярусе t расположено β_j вершин, помеченных нетерминалом A_j, $1 \leqslant j \leqslant k$.

Введем многомерные производящие функции $F_i(t, s) = \sum\limits_{\beta} p^i_\beta(t) s^\beta$, $1 \leqslant j \leqslant k$, где $s = (s_1, s_2, \dots, s_k)$ и $s^\beta = s_1^{\beta_1} s_2^{\beta_2} \dots s_k^{\beta_k}$.

На множестве деревьев вывода с корнем, помеченным нетерминалом A_i, рассмотрим случайную величину $\mu_{ij}(t)$ — число вершин на ярусе t дерева вывода, помеченных нетерминалом A_j. Математическое ожидание величины $\mu_{ij}(t)$ обозначим через $a_{ij}(t)$. Очевидно, что

$$a_{ij}(t) = \left.\frac{\partial F_i(t,s)}{\partial s_j}\right|_{s_1=s_2=\ldots=s_k=1}.$$

Отметим, что $a_{ij}(t)$ — элемент матрицы A^t [6] и $a_{ij}(t) = O(1)$ при $r = 1$, что следует из свойств неразложимой непериодической матрицы [2]. Заметим, что $F_i(1,s) = F_i(s)$ ($F_i(s)$ определено выше) и матрица первых моментов состоит из элементов $a_{ij}(1)$, которые выше обозначены через a_{ij}.

Известно необходимое и достаточное условие согласованности стохастической КС-грамматики, следующее из результатов в [6]: стохастическая КС-грамматика при отсутствии бесполезных нетерминалов (т. е. не участвующих в порождении слов языка) является согласованной тогда и только тогда, когда $r \leqslant 1$. Так как мы рассматриваем случай $r = 1$, для обеспечения согласованности грамматики будем полагать, что нет бесполезных нетерминалов.

2. Некоторые предварительные результаты

Пусть G — стохастическая КС-грамматика и A_i — некоторый нетерминальный символ. Через L_i обозначим язык, порожденный грамматикой G_i, которая получается из исходной грамматики G заменой аксиомы на нетерминал A_i. Будем считать, что аксиомой исходной грамматики является первый нетерминальный символ A_1 и $L = L_1$ для исходной грамматики G.

Через D_i обозначим множество деревьев вывода для слов из L_i и через D_i^t — множество деревьев вывода высоты t для слов из L_i. В дальнейшем будем опускать индекс i в обозначениях, если $i = 1$ и это не ведет к неопределенности.

Будем полагать, что D_i^1 не пусто при любом $i = 1, \ldots, k$. Это означает, что для любого нетерминала A_i существует правило грамматики вида $A_i \xrightarrow{p_{ij}} \beta$, где β содержит только терминальные символы. Это предположение не уменьшает общности дальнейших результатов, так как при отсутствии в грамматике бесполезных нетерминалов всегда можно перейти к эквивалентной грамматике, обладающей требуемым свойством, применяя метод укрупнения правил из [4]. При этом предположении D_i^t не пусто при любом t.

Обозначим через $Q_i(t)$ вероятность появления деревьев вывода для грамматики G_i, имеющих высоту, большую t.

Лемма 1 [6]. *При любом $i \in \{1, \dots, k\}$ справедливо равенство $Q_i(t) = \frac{2u_i}{Bt}\left(1 + \zeta_i(t)\right)$, где $\zeta_i(t) = o(1)$, константа B задается формулой*

$$B = \sum_{l,m,n} v_n u_l u_m b_{nlm}, \tag{1}$$

в которой b_{nlm} — вторые моменты, а $U = (u_1, \dots, u_k)$ и $V = (v_1, \dots, v_k)$ — соответственно правый и левый положительные собственные векторы для перронова корня r при нормировке $\sum\limits_{i=1}^{k} u_i v_i = 1$.

Доказательство леммы получается непосредственной интерпретацией результатов теории ветвящихся процессов к процессу порождения слов КС-языка.

Лемма 2. *При любом $i \in \{1, \dots, k\}$ вероятность $P\left(D_i^t\right)$ удовлетворяет соотношению*

$$P\left(D_i^t\right) = \frac{2u_i}{Bt^2}\left(1 + \phi_i(t)\right),$$

где $\phi_i(t) = o(1)$, а B и u_i имеют тот же смысл, что и в лемме 1.

Доказательство. Сначала докажем , что

$$\sum_{i=1}^{k} v_i P\left(D_i^t\right) = \frac{2}{Bt^2}\left(1 + o(1)\right).$$

Для представления функции $R_i(s) = 1 - F_i(s)$ воспользуемся разложением производящей функции $F_i(s)$ в ряд Тейлора в окрестности точки $s = (1, 1, \dots, 1)$. Поскольку $F_i(1) = \sum\limits_j p_{ij} = 1$ и, следовательно, $R_i(1) = 0$, можно записать, что

$$\begin{aligned} R_i(s) = \sum_{j=1}^{k} a_{ij}(1 - s_j) - \frac{1}{2}\sum_{j,l} b_{ijl}(1 - s_j)(1 - s_l) \\ + \frac{1}{6}\sum_{j,l,m} c_{ijlm}(\theta)(1 - s_j)(1 - s_l)(1 - s_m), \end{aligned} \tag{2}$$

где $c_{ijlm}(\theta)$ — значение третьей производной производящей функции $F_i(s)$ по переменным s_j, s_l и s_m в точке θ и $c_{ijlm}(\theta) \leqslant c_{ijlm}(1)$ при $0 \leqslant \theta \leqslant 1$.

Умножив (2) на v_i и просуммировав по i, получаем

$$\sum_{i=1}^{k} v_i R_i(s) = \sum_{i=1}^{k} v_i \sum_{j=1}^{k} a_{ij}(1-s_j) - \frac{1}{2}\sum_{i=1}^{k} v_i \sum_{j,l} b_{ijl}(1-s_j)(1-s_l)$$
$$+ \frac{1}{6}\sum_{i=1}^{k} v_i \sum_{j,l,m} c_{ijlm}(\theta)(1-s_j)(1-s_l)(1-s_m).$$

Так как $V = (v_1, \ldots, v_k)$ — левый собственный вектор, соответствующий перронову корню $r = 1$, то справедливы соотношения

$$\sum_{i=1}^{k} v_i \sum_{j=1}^{k} a_{ij}(1-s_j) = \sum_{j=1}^{k}(1-s_j)\sum_{i=1}^{k} v_i a_{ij} = \sum_{j=1}^{k} v_j(1-s_j).$$

Поэтому

$$\sum_{i=1}^{k} v_i R_i(s) = \sum_{i=1}^{k} v_i(1-s_i) - \frac{1}{2}\sum_{i=1}^{k} v_i \sum_{j,l} b_{ijl}(1-s_j)(1-s_l)$$
$$+ \frac{1}{6}\sum_{i=1}^{k} v_i \sum_{j,l,m} c_{ijlm}(\theta)(1-s_j)(1-s_l)(1-s_m). \qquad (3)$$

Положим $F(t,s) = (F_1(t,s), F_2(t,s), \ldots, F_k(t,s))$. Подставим в (3) вместо s_i функцию $F_i(t-1,s)$ и учтем, что $F_i(F(t-1,s)) = F_i(t,s)$ [6]. Тогда, применяя обозначение $R_i(t,s) = 1 - F_i(t,s)$, равенство (3) можно переписать в виде

$$\sum_{i=1}^{k} v_i R_i(t,s) = \sum_{i=1}^{k} v_i R_i(t-1,s) - \frac{1}{2}\sum_{i=1}^{k} v_i \sum_{j,l} b_{ijl} R_j(t-1,s) R_l(t-1,s)$$
$$+ \frac{1}{6}\sum_{i=1}^{k} v_i \sum_{j,l,m} c_{ijlm}(\theta) R_j(t-1,s) R_l(t-1,s) R_m(t-1,s).$$

Используя соотношения $Q_i(t) = R_i(t,0)$ [6],

$$P\left(D_i^t\right) = Q_i(t-1) - Q_i(t),$$

лемму 1 и равенство (1), получаем

$$\begin{aligned}\sum_{i=1}^{k} v_i P(D_i^t) &= \frac{1}{2}\sum_{i=1}^{k} v_i \sum_{jl} b_{ijl} u_j u_l \frac{4}{B^2 t^2}(1+o(1)) + O\left(\frac{1}{t^3}\right) \\ &= \frac{2}{Bt^2}(1+o(1)).\end{aligned} \qquad (4)$$

Для завершения доказательства леммы остается показать, что

$$\lim_{t\to\infty} \frac{P(D_i^t)}{\sum\limits_{j=1}^{k} v_j P(D_j^t)} = u_i.$$

Подставим в (2) величину $1 - Q_j(t-1)$ вместо s_j $(1 \leqslant j \leqslant k)$. Так как $1 - Q_j(t-1) = F_j(t-1, 0)$, то

$$R_i(1 - Q(t-1)) = 1 - F_i(F(t-1,0)) = 1 - F_i(t,0) = Q_i(t)$$

(здесь через $Q(t-1)$ обозначен вектор $(Q_1(t-1), \ldots, Q_k(t-1))$).

После подстановки уравнение (2) примет следующий вид

$$Q_i(t) = \sum_{j=1}^{k} a_{ij} Q_j(t-1) - \frac{1}{2} \sum_{j,l} b_{ijl} Q_j(t-1) Q_l(t-1)$$
$$+ \tfrac{1}{6} \sum_{j,l,m} c_{ijlm}(\theta) Q_j(t-1) Q_l(t-1) Q_m(t-1).$$

Вычитая это уравнение из аналогичного уравнения для $Q_i(t-1)$, после несложных преобразований получим уравнение

$$P(D_i^t) = \sum_{j=1}^{k} a_{ij} P(D_j^{t-1}) - \frac{1}{2} \sum_{j,l} b_{ijl} (Q_j(t-2) P(D_l^{t-1})$$
$$+ Q_l(t-1) P(D_j^{t-1})) + \frac{1}{6} \sum_{j,l,m} c_{ijlm}(\theta) (Q_j(t-2) Q_l(t-2) P(D_m^{t-1})$$
$$+ Q_j(t-2) Q_m(t-1) P(D_l^{t-1}) + Q_l(t-1) Q_m(t-1) P(D_j^{t-1})).$$

Рассматривая $P(D^t)$ как вектор $\left(P(D_1^t), \ldots, P(D_k^t)\right)$, полученное уравнение можно записать в матричном виде

$$P(D^t) = (A - E_t) P(D^{t-1}),$$

где каждый элемент матрицы E_t не превосходит $O\left(\frac{1}{t}\right)$. Раскрывая $P(D^{t-1})$, представим $P(D^t)$ в виде:

$$P(D^t) = (A - E_t)(A - E_{t-1}) \ldots (A - E_1) P(D^1).$$

Положим $B_t = (A - E_t)(A - E_{t-1}) \ldots (A - E_1)$. В [6] доказано следующее

Утверждение. *Пусть A — неразложимая непериодическая матрица, перронов корень которой равен 1, и $E_1, E_2, \ldots, E_t$ — последовательность таких матриц, что $0 \leqslant E_n \leqslant A$ $(1 \leqslant n \leqslant t)$ и $\lim E_t = 0$ при $t \to \infty$. Тогда*

$$\lim_{t\to\infty} \frac{B_t X}{V B_t X} = U \tag{5}$$

для любого вектора $X \geqslant 0$, удовлетворяющего условию $B_n X \neq 0$ при любом $n \geqslant 1$. (U и V — соответственно правый и левый собственные векторы для $r = 1$).

Применяя равенство (5) к вектору $P(D^1)$, получаем

$$\lim_{t\to\infty} \frac{P(D_i^t)}{\sum\limits_{j=1}^{k} v_j P(D_j^t)} = u_i.$$

Из этого утверждения и равенства (4) следует утверждение леммы 2.

Рассмотрим случайную величину $\xi_j^m(\tau) = \dfrac{2\mu_{mj}(\tau)}{B\tau v_j}$, $j \in \{1, \ldots, k\}$, где $\mu_{mj}(\tau)$ — число вершин на ярусе τ дерева вывода из D_m^t, помеченных нетерминалом A_j. В дальнейшем через $\xi^m(\tau)$ будем обозначать случайный вектор $(\xi_1^m(\tau), \ldots, \xi_k^m(\tau))$ и через $\mu_m(\tau)$ — случайный вектор $(\mu_{m1}(\tau), \ldots, \mu_{mk}(\tau))$.

Лемма 3 [6]. *Последовательность случайных векторов $\xi^m(\tau)=(\xi_1^m(\tau), \ldots, \xi_k^m(\tau))$ при условии $\xi^m(\tau) \neq 0$ сходится по распределению при $\tau \to \infty$ к случайному вектору $\xi = (\xi_1, \ldots, \xi_k)$, не зависящему от m. При этом $\xi_1 = \xi_2 = \ldots = \xi_k$ с вероятностью 1 и*

$$P\left(\xi_1 \leqslant y\right) = 1 - e^{-y}\ , y \geqslant 0.$$

Пусть $\xi = (\xi_1, \ldots, \xi_k)$ — случайный вектор из леммы 3 и n — неотрицательное целое число. Возьмем ε из интервала $(0, 1)$. Определим множества полуинтервалов $B_{nj} = (n\varepsilon, (n+1)\varepsilon]$ $(1 \leqslant j \leqslant k)$. Пусть $M_n = B_{n1} \times B_{n2} \times \ldots \times B_{nk}$ — декартово произведение. Положим $M = \bigcup\limits_{n=0}^{\infty} M_n$. Множество M таково, что все точки с равными координатами $(\xi_1, \ldots, \xi_k)$, т. е. $\xi_1 = \ldots = \xi_k$, кроме точки $(0, 0, \ldots, 0)$, принадлежат M. Очевидно, что $M_n \cap M_m = \emptyset$ при $m \neq n$.

Из леммы 3 следуют равенства $P(\xi \in M_n) = e^{-n\varepsilon}\left(1 - e^{\varepsilon}\right)$ и $P(\xi \in \Gamma(M_n)) = 0$, где $\Gamma(M_n)$ — граница множества M_n. Поэтому из определения сходимости по распределению [8] получаем

$$P\left(\xi^m(\tau) \in M_n | \xi^m(\tau) \neq 0\right) \to \left(1 - e^{\varepsilon}\right) e^{-n\varepsilon}.$$

Аналогично устанавливается соотношение $P\left(\xi^m(\tau) \in M | \xi^m(\tau) \neq 0\right) \to 1$.

Определим множества полуинтервалов

$$B^*_{nj} = \left(\frac{\varepsilon n B \tau v_j}{2}, \frac{\varepsilon (n+1) B \tau v_j}{2} \right], \ 1 \leqslant j \leqslant k.$$

Положим $M^*_n = B^*_{n1} \times B^*_{n2} \times \ldots \times B^*_{nk}$ и $M^* = \cup_{n=0}^{\infty} M^*_n$. Очевидно, что $P\left(\mu_m(\tau) \in M^*_n | \mu_m(\tau) \neq 0\right) = P\left(\xi^m(\tau) \in M_n | \xi^m(\tau) \neq 0\right)$. Поэтому справедливо

Следствие 1. *При* $\tau \to \infty$

1) $P(\mu_m(\tau) \in M^*_n | \mu_m(\tau) \neq 0) = \left(1 - e^{-\varepsilon}\right) e^{-n\varepsilon} + \delta_n$, где $\delta_n \to 0$,

2) $P(\mu_m(\tau) \in M^* | \mu_m(\tau) \neq 0) = 1 + \Delta$, *где* $\Delta = \sum\limits_{n=0}^{\infty} \delta_n \to 0$.

Положим

$$R_X(n) = \prod_{j=1}^{k} (1 - Q_j(n))^{x_j} - \prod_{j=1}^{k} (1 - Q_j(n-1))^{x_j}, \quad X = (x_1, \ldots, x_k).$$

Лемма 4. *Пусть* $X = (x_1, \ldots, x_k)$ *— неотрицательный целочисленный вектор и* n *— натуральное число. Тогда при* $n \to \infty$

$$R_X(n) = \prod_{j=1}^{k} \left(1 - Q_j(n) \left(1 + \psi_j(n)\right)\right)^{x_j} \sum_{l=1}^{k} x_l P(D^n_l)(1 + \gamma_l(n)),$$

где $0 \leqslant \psi_j(n) \leqslant \dfrac{2}{n}$ *и* $0 \leqslant \gamma_l(n) \leqslant \dfrac{4u_l}{Bn}$ $(j, l \in \{1, \ldots, k\})$.

Доказательство. Индукцией по k легко показать, что

$$\sum_{l=1}^{k} \frac{x_l P(D^n_l)}{1 - Q_l(n-1)} \prod_{j=1}^{k} (1 - Q_j(n-1))^{x_j} \leqslant R_X(n)$$

$$\leqslant \sum_{l=1}^{k} \frac{x_l P(D^n_l)}{1 - Q_l(n)} \prod_{j=1}^{k} (1 - Q_j(n))^{x_j}. \qquad (6)$$

Отсюда следует, что

$$R_X(n) = \prod_{j=1}^{k} \left(1 - Q_j(n) \left(1 + \psi_j(n)\right)\right)^{x_j} \sum_{l=1}^{k} x_l P(D^n_l)(1 + \gamma_l(n)),$$

где $Q_j(n) \leqslant Q_j(n)(1+\psi_j(n)) \leqslant Q_j(n-1)$ и

$$\frac{1}{1-Q_l(n)} \leqslant (1+\gamma_l(n)) \leqslant \frac{1}{1-Q_l(n-1)}.$$

Используя лемму 1, при $n \to \infty$ получаем ограничения для функций $\psi_j(n)$ и $\gamma_l(n)$: $0 \leqslant \psi_j(n)) \leqslant \dfrac{2}{n}$ и $0 \leqslant \gamma_l(n) \leqslant \dfrac{4u_l}{Bn}$. Лемма 4 доказана.

Лемма 5. *Пусть* $x \geqslant 0$. *Тогда при любом натуральном* n *и* $j = 1, \ldots, k$ *выполняются неравенства*

$$x\,(1-Q_j(n))^x \leqslant \frac{1}{Q_j(n)} \text{ и } x^2\,(1-Q_j(n))^x \leqslant \frac{4}{Q_j^2(n)}.$$

Доказательство. Используя две первые производные функции $f(x) = x(1-Q_j(n))^x$, убеждаемся в том, что при $x \geqslant 0$ она сначала возрастает, затем убывает и принимает максимальное значение при $x = x_0 = -\dfrac{1}{\ln(1-Q_j(n))}$. Так как

$$-\frac{1}{\ln(1-Q_j(n))} = \frac{1}{\sum\limits_{s=1}^{\infty} Q_j^s(n)/s} = x_0 \leqslant \frac{1}{Q_j(n)},$$

то при любом $x \geqslant 0$

$$x\,(1-Q_j(n))^x \leqslant x_0\,(1-Q_j(n))^{x_0} \leqslant x_0 \leqslant \frac{1}{Q_j(n)}.$$

Аналогично доказывается второе утверждение.

3. Закономерности в деревьях вывода для критического случая

Пусть G — стохастическая КС-грамматика, матрица первых моментов которой неразложима, непериодична и ее перронов корень равен 1. Через $M_i(t,\tau)$ обозначим условное математическое ожидание числа вершин, помеченных нетерминалом A_i, в деревьях вывода высоты t на ярусе τ, $1 \leqslant i \leqslant k$.

Теорема 1. *Пусть* D_1^t — *множество деревьев вывода высоты* t *для слов языка, порождаемого стохастической КС-грамматикой с неразложимой и непериодической матрицей первых моментов, перронов корень* *которой равен единице. Тогда для любого* $\varepsilon \in (0,1)$ *при* $t \to \infty$ *и* $t\sqrt{\varepsilon} \leqslant \tau \leqslant t(1-\sqrt{\varepsilon})$ *выполняется равенство*

$$M_i(t,\tau) = \frac{v_i B \tau (t-\tau)}{t}\,(1+\chi_i(t,\tau,\varepsilon)) \;\; (i=1,\ldots,k), \tag{7}$$

где $|\chi_i(t,\tau,\varepsilon)| \leqslant c_0\varepsilon$ *и* c_0 *— некоторая константа, не зависящая от* t *и* τ, v_i *есть* i*-я компонента левого собственного вектора для перронова корня и* B *— константа, определяемая формулой* (1).

Доказательство. Будем полагать, что аксиомой исходной грамматики является нетерминал A_1. Величину $M_i(t,\tau)$ можно записать в виде:

$$M_i(t,\tau) = \frac{1}{P(D_1^t)} \sum_{d \in D_1^t} p(d) z_i(d,\tau),$$

где $z_i(d,\tau)$ — число вершин на ярусе τ дерева d, помеченных нетерминалом A_i, $p(d)$ — вероятность появления дерева d в исходной грамматике.

Рассмотрим неотрицательный целочисленный вектор $X = (x_1,\ldots,x_k)$. Используя X, можно записать

$$M_i(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} \Delta_X,$$

где Δ_X — вклад в математическое ожидание тех деревьев вывода из D_1^t, которые на ярусе τ содержат x_j вершин, помеченных нетерминалом A_j, $1 \leqslant j \leqslant k$. Множество таких деревьев обозначим через $D_X^t(\tau)$.

Пусть $d \in D_X^t(\tau)$. Выделим в d поддерево d_0 и последовательность поддеревьев $d_1, d_2, \ldots d_m$, где $m = \sum\limits_{l=1}^{k} x_l$. Поддерево d_0 получено из d удалением всех вершин на ярусах $\tau+1$, $\tau+2, \ldots, t$ и инцидентных им дуг. Последовательность d_1, $d_2, \ldots d_m$ образуют все поддеревья, корни которых расположены на ярусе τ дерева d. При этом корни поддеревьев d_1, $d_2, \ldots d_m$ расположены в дереве d последовательно в порядке обхода вершин яруса τ слева направо, и каждое дерево d_l $(l=1,\ldots,m)$ содержит все дуги и вершины дерева d, лежащие на путях от корня d_l к листьям дерева d.

Выделим в $D_X^t(\tau)$ множество деревьев, имеющих в качестве поддерева d_0 одно и то же дерево. Это множество обозначим через D_0. Нетрудно понять, что

$$P\left(D_0\right) = p(d_0)\left(\prod_{l=1}^{k}(1-Q_l(t-\tau))^{x_l} - \prod_{l=1}^{k}(1-Q_l(t-\tau-1))^{x_l}\right), \tag{8}$$

поскольку $(1 - Q_l(n))$ — вероятность появления деревьев вывода высоты не более n с корнем, помеченным нетерминалом A_l.

Положим

$$\delta_1(X) := \prod_{l=1}^{k}(1 - Q_l(t-\tau))^{x_l} \text{ и } \delta_2(X) := \prod_{l=1}^{k}(1 - Q_l(t-\tau-1))^{x_l}.$$

В (8) величина $p(d_0)\delta_1(X)$ есть вероятность появления таких деревьев высоты не более t, определяемых поддеревом d_0, что каждое поддерево с корнем на ярусе τ имеет высоту, не превосходящую $t-\tau$. Вторая величина $p(d_0)\delta_2(X)$ есть вероятность появления деревьев высоты не более $t-1$, определяемых поддеревом d_0.

Разность $p(d_0)\delta_1(X) - p(d_0)\delta_2(X)$ равна, очевидно, вероятности появления деревьев высоты t, определяемых деревом d_0, и значение $\delta_1(X) - \delta_2(X)$ не зависит от порядка следования вершин на ярусе τ, помеченных нетерминалами. Поэтому

$$P(D_X^t(\tau)) = (\delta_1(X) - \delta_2(X)) \sum_{d_0} p(d_0),$$

где суммирование осуществляется по всем поддеревьям d_0 деревьев вывода из $D_X^t(\tau)$.

Для каждой вершины, помеченной некоторым нетерминалом A_l, вероятность появления деревьев с корнем в этой вершине и листьями, помеченными только терминалами, равна $P(D_l)$. Легко показать, что $P(D_l) = 1$ при любом l ввиду согласованности исходной грамматики G. Поэтому

$$\sum_{d_0} p(d_0) = \sum_{d_0} p(d_0) P(D_1)^{x_1} P(D_2)^{x_2} \dots P(D_k)^{x_k} = \sum_{d \in D_X(\tau)} p(d),$$

где $D_X(\tau)$ — множество деревьев из D_1, имеющих x_j вершин на ярусе τ, помеченных нетерминалами A_j $(1 \leqslant j \leqslant k)$.

Положим $P_X(\tau) := \sum\limits_{d \in D_X(\tau)} p(d)$. Тогда

$$M_i(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} P_X(\tau)\,(\delta_1(X) - \delta_2(X))\, x_i.$$

В обозначениях раздела 2 разность $\delta_1(X) - \delta_2(X)$ есть $R_X(t-\tau)$.

Пусть M^* — множество вещественных неотрицательных векторов, определенное в разделе 2. (Напомним, что любой вектор $X \in M^*$ близок к вектору вида bV, где V — левый собственный вектор для перронова корня и $b > 0$.) Тогда

$$M_i(t,\tau) = \frac{1}{P(D_1^t)}\left(\sum_{X\in M^*} P_X(\tau)R_X(t-\tau)x_i + \sum_{X\in\overline{M}^*} P_X(\tau)R_X(t-\tau)x_i\right),$$

где $\overline{M}^*$ — дополнение множества M^* до множества всех вещественных неотрицательных векторов.

Пусть

$$S_1 := \sum_{X\in M^*} P_X(\tau)R_X(t-\tau)x_i \tag{9}$$

и

$$S_2 := \sum_{X\in\overline{M}^*} P_X(\tau)R_X(t-\tau)x_i. \tag{10}$$

Отдельно вычислим эти суммы. Так как $M^* = \bigcup\limits_{n=0}^{\infty} M_n^*$, то

$$S_1 = \sum_{n=0}^{\infty}\sum_{X\in M_n^*} P_X(\tau)R_X(t-\tau)x_i.$$

Пусть $X \in M_n^*$. В M_n^* представим x_l в виде:

$$x_l = \frac{n\varepsilon B\tau v_l}{2} + \Delta(x_l), \quad \text{где} \quad 0 < \Delta(x_l) \leqslant \frac{\varepsilon B\tau v_l}{2}.$$

К $P_X(\tau)$ применим следствие 1 при $m = 1$. Так как

$$\sum_{X\in M_n^*} P_X(\tau) = P\left(\mu_1(\tau) \in M_n^* | \mu_1(\tau) \neq 0\right) P(\mu_1(\tau) \neq 0),$$

а $P(\mu_1(\tau) \neq 0) = Q_1(\tau)$, то

$$\sum_{X\in M_n^*} P_X(\tau) = \left((1-e^{-\varepsilon})e^{-n\varepsilon} + \delta_n\right) Q_1(\tau). \tag{11}$$

Применяя лемму 4 к $R_X(t-\tau)$, для S_1 получаем верхнюю оценку

$$S_1 \leqslant S_1^{\text{в}} = (1 + \max_j\{\gamma_j^{\text{в}}(t-\tau)\})Q_1(\tau)\sum_{j=1}^{k} P(D_j^{t-\tau})\sum_{n=0}^{\infty}\left((1-e^{-\varepsilon})e^{-n\varepsilon} + \delta_n\right)$$

$$\times \left(\frac{(n+1)^2\varepsilon^2 B^2\tau^2 v_i v_j}{4}\right)\prod_{l=1}^{k}(1-Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{n\varepsilon B\tau v_l/2}$$

$$= (1+\max_j\{\gamma_j^{\text{в}}(t-\tau)\})Q_1(\tau)\sum_{j=1}^{k}P(D_j^{t-\tau})\sum_{n=0}^{\infty}\big((1-e^{-\varepsilon})e^{-n\varepsilon}+\delta_n\big)$$

$$\times \left(\frac{n^2\varepsilon^2 B^2\tau^2 v_i v_j}{4}\right)\prod_{l=1}^{k}(1-Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{n\varepsilon B\tau v_l/2}$$
$$\times (1+\Delta(n))\,, \tag{12}$$

где $\gamma_j^{\text{в}}(t-\tau) = \dfrac{4u_l}{B(t-\tau)}$, $\psi_l^{\text{в}}(t-\tau) = 0$ и $\Delta(n) = \left(\dfrac{n+1}{n}\right)^2 - 1 \leqslant \dfrac{3}{n}$ при $n > 0$.

Аналогично получаем

$$S_1 \geqslant S_1^{\text{н}} = Q_1(\tau)\sum_{j=1}^{k}P(D_j^{t-\tau})\sum_{n=0}^{\infty}\big((1-e^{-\varepsilon})e^{-n\varepsilon}+\delta_n\big)$$

$$\times \frac{n^2\varepsilon^2 B^2\tau^2 v_i v_j}{4}\prod_{l=1}^{k}(1-Q_l(t-\tau)(1+\psi_l^{\text{н}}(t-\tau)))^{(n+1)\varepsilon B\tau v_l/2}\,. \tag{13}$$

где $\psi_l^{\text{н}}(t-\tau) = \dfrac{2}{t-\tau}$.

Вычислим $S_1^{\text{в}}$ и $S_1^{\text{н}}$. Представим $S_1^{\text{в}}$ в следующем виде

$$S_1^{\text{в}}=(1+\max_j \gamma_j^{\text{в}}(t-\tau))\frac{B^2 v_i\varepsilon^2\tau^2}{4}Q_1(\tau)\sum_{j=1}^{k}v_j P(D_j^{t-\tau})\,(S_{11}+S_{12}+S_{13})\,, \tag{14}$$

где

$$S_{11}=(1-e^{-\varepsilon})\sum_{n=0}^{\infty}e^{-n\varepsilon}n^2\left(\prod_{l=1}^{k}(1-Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B\tau v_l/2}\right)^n, \tag{15}$$

$$S_{12} = (1-e^{-\varepsilon})\sum_{n=0}^{\infty}e^{-n\varepsilon}n^2\left(\prod_{l=1}^{k}(1-Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B\tau v_l/2}\right)^n$$
$$\times\Delta(n), \tag{16}$$

и

$$S_{13}=\sum_{n=0}^{\infty}\delta_n n^2(1+\Delta(n))\left(\prod_{l=1}^{k}(1-Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B\tau v_l/2}\right)^n. \tag{17}$$

Оценим сверху слагаемые S_{11}, S_{12} и S_{13}.

Для оценки S_{11} воспользуемся равенством

$$\sum_{n=1}^{\infty} n^2 x^n = \frac{x(1+x)}{(1-x)^3},$$

которое справедливо при любом x, $0 \leqslant x < 1$.

Положим

$$x = e^{-\varepsilon} \prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B \tau v_l/2}.$$

Тогда

$$S_{11} = (1-e^{-\varepsilon}) \left(\prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B \tau v_l/2} \right)$$
$$\times \frac{e^{-\varepsilon} \left(1 + e^{-\varepsilon} \prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B \tau v_l/2} \right)}{\left(1 - e^{-\varepsilon} \prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{v_l \varepsilon B \tau/2} \right)^3}. \quad (18)$$

Будем рассматривать значения τ, удовлетворяющие условию

$$t\sqrt[4]{\varepsilon} \leqslant \tau \leqslant t(1-\sqrt[4]{\varepsilon}). \quad (19)$$

Из неравенств (19) следует, что ярус τ находится на достаточно большом удалении от корня дерева вывода и от последнего яруса t.

Аппроксимируем $e^{-\varepsilon}$ с помощью разложения в ряд Тейлора:

$$e^{-\varepsilon} = 1 - \varepsilon + \frac{\varepsilon^2}{2}(1 + O(\varepsilon)). \quad (20)$$

Здесь и далее запись $O(f(\varepsilon))$ применяется для обозначения величины, не превосходящей по модулю $cf(\varepsilon)$, где $c > 0$ — некоторая константа.

Для оценки сверху выражения $\prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l^{\text{в}}(t-\tau)))^{\varepsilon B \tau v_l/2}$ воспользуемся следующим равенством (см. [7]):

$$(1-y_1)^{n_1} \ldots (1-y_k)^{n_k} = 1 - \sum_{i=1}^{k} n_i y_i + R_2, \quad (21)$$

где $0 \leqslant R_2 \leqslant \sum_{i<j} n_i n_j y_i y_j + \sum_{i=1}^{k} \binom{n_i}{2} y_i^2.$

Используя (21), нормировку $\sum_{l=1}^{k} u_l v_l = 1$, а также неравенства

$$\left\lfloor \frac{\varepsilon B\tau v_l}{2} \right\rfloor \leqslant \frac{\varepsilon B\tau v_l}{2} < \left\lfloor \frac{\varepsilon B\tau v_l}{2} \right\rfloor + 1,$$

можно записать

$$\prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l(t-\tau)))^{\varepsilon B\tau v_l/2} = 1 - \sum_{i=1}^{k} \frac{\varepsilon\tau v_l u_l (1+\eta_l(t-\tau))}{t-\tau} + R_2$$
$$= 1 - \frac{\varepsilon\tau(1+\eta(t-\tau))}{t-\tau} + R_2,$$

где $\eta_l(t-\tau) = o(1)$ при $l = 1, \ldots, k$, $\eta(t-\tau) = o(1)$ при $t - \tau \to \infty$ и $0 \leqslant R_2 \leqslant c\varepsilon^2\tau^2(t-\tau)^{-2}$ при некоторой константе c. С учетом (19) для R_2 имеем $R_2 \leqslant c\varepsilon^{\frac{3}{2}}$. Поэтому

$$\prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l(t-\tau)))^{\varepsilon B\tau v_l/2} = 1 - \frac{\varepsilon\tau(1+\eta(t-\tau))}{t-\tau} + O\left(\varepsilon^{\frac{3}{2}}\right). \tag{22}$$

Будем также использовать соотношение

$$\prod_{l=1}^{k} (1 - Q_l(t-\tau)(1+\psi_l(t-\tau)))^{\varepsilon B\tau v_l/2} = 1 + O\left(\varepsilon^{\frac{3}{4}}\right). \tag{23}$$

Из (15), (18), (20) и (23) после несложных преобразований с учетом соотношений (19) получаем

$$S_{11} = \frac{2 + O(\varepsilon^{\frac{3}{4}})}{\varepsilon^2 \left(\frac{t}{t-\tau} + O(\varepsilon) + O(\varepsilon^{-\frac{1}{4}})\eta(t-\tau) + O\left(\varepsilon^{\frac{1}{2}}\right) \right)^3}.$$

Так как для любого $\varepsilon > 0$ выполняется неравенство $\eta(t-\tau) \leqslant \varepsilon$, начиная с некоторого $t-\tau$, то $O(\varepsilon^{-\frac{1}{4}})\eta(t-\tau) = O(\varepsilon^{\frac{3}{4}})$. Поэтому

$$S_{11} = \frac{2(1+O(\varepsilon^{\frac{3}{4}}))(t-\tau)^3}{\varepsilon^2 t^3 (1+O(\sqrt{\varepsilon}))} = \frac{2(t-\tau)^3}{\varepsilon^2 t^3}(1+O(\sqrt{\varepsilon})). \tag{24}$$

Так как $\Delta(n) \leqslant \frac{3}{n}$, то из (16) следует, что

$$S_{12} \leqslant 3(1-e^{-\varepsilon}) \sum_{n=0}^{\infty} e^{-n\varepsilon} n \left(\prod_{l=1}^{k} (1-Q_l(t-\tau))^{\varepsilon B \tau v_l/2} \right)^n$$

$$\leqslant 3(1-e^{-\varepsilon}) \sum_{n=0}^{\infty} e^{-n\varepsilon} n \left((1-Q_1(t-\tau))^{\varepsilon B \tau v_1/2} \right)^n .$$

Воспользовавшись леммой 5, получим

$$S_{12} \leqslant 3(1-e^{\varepsilon}) \sum_{n=0}^{\infty} e^{-n\varepsilon} \frac{1}{Q_1(t-\tau)} \frac{2}{\varepsilon B \tau v_1} \leqslant \frac{4(t-\tau)}{\varepsilon \tau u_1 v_1} \qquad (25)$$

и

$$S_{13} \leqslant \sum_{n=0}^{\infty} |\delta_n| c_2 \frac{(t-\tau)^2}{\varepsilon^2 \tau^2}$$

при некоторой константе $c_2 > 0$.

Множество M^* разобьем на два подмножества M^{*+} и M^{*-} : множество M_n^* отнесем к M^{*+}, если $\delta_n \geqslant 0$, и M_n^* отнесем к M^{*-}, если $\delta_n < 0$. Тогда можно записать следующее неравенство:

$$S_{13} \leqslant c_2 \frac{(t-\tau)^2}{\varepsilon^2 \tau^2} \left(\sum_{M^{*+}} \delta_n + |\sum_{M^{*-}} \delta_n| \right).$$

Пусть $\delta := \sum\limits_{M^{*+}} \delta_n$ и $\delta^- := \sum\limits_{M^{*-}} \delta_n$. Так как имеет место сходимость по распределению, то $\delta^+ \to 0$ и $\delta^- \to 0$ при $\tau \to \infty$. Положим $\delta = \delta^+ + |\delta^-|$. Тогда

$$S_{13} \leqslant c_2 \delta \frac{(t-\tau)^2}{\varepsilon^2 \tau^2} \ \delta \to 0 \text{ при } \ \tau \to \infty. \qquad (26)$$

Пользуясь (24)–(26), получаем

$$S_{11} + S_{12} + S_{13} = \frac{2(t-\tau)^3}{\varepsilon^2 t^3} \left(1 + O(\sqrt{\varepsilon}) + O\left(\frac{\varepsilon t^3}{\tau (t-\tau)^2} \right) + O\left(\frac{\delta t^3}{\tau^2 (t-\tau)} \right) \right).$$

Поскольку либо $\dfrac{\tau}{t} \geqslant 1/2$, либо $\dfrac{t-\tau}{t} \geqslant 1/2$ и справедливо (19), имеем

$$O\left(\frac{\varepsilon t^3}{\tau (t-\tau)^2} \right) \leqslant O(\sqrt{\varepsilon}) \ \ O\left(\frac{\delta t^3}{\tau^2 (t-\tau)} \right) \leqslant O\left(\frac{\delta}{\sqrt{\varepsilon}} \right).$$

Так как $\delta \to 0$ при $\tau \to \infty$, то найдется такое t_0, что при $t \geqslant t_0$ справедливо неравенство $\delta \leqslant \varepsilon$. Тогда

$$S_{11} + S_{12} + S_{13} = \frac{2(t-\tau)^3}{\varepsilon^2 t^3}\left(1 + O(\sqrt{\varepsilon})\right). \qquad (27)$$

Из (14) и (27) получаем

$$S_1^{\text{в}} = (1+\max_j\{\gamma_j(t-\tau)\})Q_1(\tau)\sum_{j=1}^{k} P\left(D_j^{t-\tau}\right)\frac{B^2 v_i v_j \tau^2 (t-\tau)^3}{2t^3}(1+O(\sqrt{\varepsilon})).$$

Воспользовавшись леммами 1 и 2 для представлений $Q_1(\tau)$ и $P\left(D_j^{t-\tau}\right)$ и оценкой для $\gamma_j^{\text{в}}(t-\tau)$, следующей из леммы 4, получаем

$$S_1^{\text{в}} = \left(1 + O\left(\frac{1}{n}\right)\right)\frac{2u_1}{B\tau}(1+\zeta_1(\tau))\sum_{j=1}^{k}\frac{2u_j}{B(t-\tau)^2}(1+\phi_j(t-\tau))$$

$$\times\frac{B^2 v_i v_j \tau^2 (t-\tau)^3}{2t^3}\left(1+O(\sqrt{\varepsilon})\right) = 2u_1 v_i \frac{\tau(t-\tau)}{t^3}\left(1+O(\sqrt{\varepsilon})\right).$$

Аналогично найдем нижнюю оценку для $S_1^{\text{н}}$. Для этого достаточно провести те же самые преобразования, что и для $S_1^{\text{в}}$, заменив $\gamma_j^{\text{в}}(t-\tau)$ на $\gamma_j^{\text{н}}(t-\tau)=0$, $\psi_l^{\text{в}}(t-\tau)$ на $\psi_l^{\text{н}}(t-\tau)$, положив $\Delta = 0$ и воспользовавшись (27). В результате получаем

$$S_1^{\text{н}} = 2u_1 v_i \frac{\tau(t-\tau)}{t^3}\left(1+O(\sqrt{\varepsilon})\right).$$

Следовательно,

$$S_1 = 2u_1 v_i \frac{\tau(t-\tau)}{t^3}\left(1+O(\sqrt{\varepsilon})\right).$$

Наконец, оценим сверху величину S_2. Применяя лемму 4 к $R_X(t-\tau)$ и лемму 5, получаем

$$S_2 \leqslant \sum_{X\in\overline{M}^*} P_X(\tau)x_i \sum_{j=1}^{k} P(D_j^{t-\tau})(1+\gamma_j(t-\tau))x_j \prod_{l=1}^{k}(1-Q_j(t-\tau))^{x_l}$$

$$\leqslant c(t-\tau)^2 \sum_{X\in\overline{M}^*} P_X(\tau)\sum_{j=1}^{k} P(D_j^{t-\tau}) = O(1)\sum_{X\in\overline{M}^*} P_X(\tau) \leqslant O(1)\delta Q_1(\tau)$$

(здесь c — некоторая константа).

Если $\delta \leqslant \varepsilon$, то

$$\sum_{X \in \overline{M}^*} P_X(\tau) \leqslant \varepsilon Q_1(\tau) = O\left(\frac{\varepsilon}{\tau}\right).$$

Таким образом,

$$\begin{aligned} S_1 + S_2 &= \frac{2u_1 v_i \tau(t-\tau)}{t^3}\left(1 + O(\sqrt{\varepsilon})\right) + O\left(\frac{\varepsilon}{\tau}\right) \\ &= \frac{2u_1 v_i \tau(t-\tau)}{t^3}\left(1 + O(\sqrt{\varepsilon}) + O\left(\frac{\varepsilon t^3}{\tau^2(t-\tau)}\right)\right). \end{aligned}$$

При выполнении (19) очевидно, что

$$O\left(\frac{\varepsilon t^3}{\tau^2(t-\tau)}\right) = O\left(\sqrt{\varepsilon}\right).$$

Поэтому

$$M_i(t,\tau) = \frac{1}{P(D_1^t)}(S_1 + S_2) = \frac{v_i B \tau(t-\tau)}{t}\left(1 + O(\sqrt{\varepsilon})\right)$$

при любом ε из интервала $(0,1)$, $t \to \infty$ и $t\sqrt[4]{\varepsilon} \leqslant \tau \leqslant t(1 - \sqrt[4]{\varepsilon})$.

В качестве нового значения ε возьмем значение ε^2. Тогда будут выполняться неравенства $t\sqrt{\varepsilon} \leqslant \tau \leqslant t(1-\sqrt{\varepsilon})$ и формула для $M_i(t,\tau)$ примет вид

$$M_i(t,\tau) = \frac{v_i B \tau(t-\tau)}{t}\left(1 + \chi_i(t,\tau,\varepsilon)\right),$$

где $\chi_i(t,\tau,\varepsilon) \leqslant c_0\varepsilon$.

Так как при получении оценки $O(\varepsilon)$ для $\chi_i(t,\tau,\varepsilon)$ использовалось конечное число раз бесконечно малые величины, зависящие от t, τ и $t-\tau$, то существует такая константа $c_0 > 0$, не зависящая от t и τ, что $|\chi_i(t,\tau,\varepsilon)| \leqslant c_0\varepsilon$. Теорема 1 доказана.

Обозначим через $M_{ij}(t,\tau)$ условное математическое ожидание числа применений правила r_{ij} на ярусе τ в деревьях вывода из D_1^t.

Теорема 2. *Пусть D_1^t — множество деревьев вывода высоты t для слов языка, порождаемого стохастической КС-грамматикой с неразложимой и непериодической матрицей первых моментов, перронов корень которой равен единице. Тогда для любого ε из интервала $(0,1)$ при $t \to \infty$ и $t\sqrt{\varepsilon} \leqslant \tau \leqslant t(1-\sqrt{\varepsilon})$ выполняется равенство*

$$M_{ij}(t,\tau) = \frac{p_{ij} v_i B \tau(t-\tau)}{t}\left(1 + \chi_{ij}(t,\tau,\varepsilon)\right) \ (i = 1,\ldots,k;\ \ j = 1,\ \ldots,n_i),$$

где $|\chi_{ij}(t,\tau,\varepsilon)| \leqslant c_0\varepsilon$, c_0 *— некоторая константа, не зависящая от* t *и* τ, *а* p_{ij} *есть вероятность применения правила* r_{ij} *в исходной грамматике.*

Доказательство. Величину $M_{ij}(t,\tau)$ можно записать в виде:

$$M_{ij}(t,\tau) = \frac{1}{P(D_1^t)} \sum_{d \in D_1^t} p(d) z_{ij}(d,\tau),$$

где $z_{ij}(d,\tau)$ — число вершин на ярусе τ дерева d, помеченных нетерминалом A_i, к которым применено правило r_{ij}, и $p(d)$ — вероятность появления дерева d в исходной грамматике.

Пусть $X = (x_1, \ldots, x_k)$ — неотрицательный целочисленный вектор. Тогда

$$M_{ij}(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} \sum_{d \in D_X^t(\tau)} p(d) z_{ij}(d,\tau).$$

Здесь $D_X^t(\tau)$ — множество деревьев вывода из D_1^t, в каждом из которых на ярусе τ содержится x_j вершин, помеченных нетерминалом A_j, $1 \leqslant j \leqslant k$. Представим $z_{ij}(d,\tau)$ в виде суммы случайных величин $I_1 + I_2 + \ldots + I_{x_i}$, где $I_m = 1$, если среди вершин, помеченных нетерминалом A_m на ярусе τ, к m-й по порядку вершине применено правило r_{ij}, и $I_m = 0$ в противном случае ($m = 1, 2, \ldots, x_i$). Тогда

$$M_{ij}(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} \sum_{d \in D_X^t(\tau)} p(d)(I_1 + I_2 + \ldots + I_{x_i}).$$

Очевидно, случайные величины I_m ($m = 1, 2, \ldots, x_i$) одинаково распределены на $D_X^t(\tau)$. Поэтому

$$M_{ij}(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} P\left(D_{X,1}^t(\tau)\right) x_i,$$

где $P(D_{X,1}^t(\tau))$ — суммарная вероятность тех деревьев из $D_X^t(\tau)$, в которых правило r_{ij} применено к первой по порядку вершине на ярусе τ, помеченной нетерминалом A_i.

Подсчитаем вероятность

$$P\left(D_{X,1}^t(\tau)\right) = p_{ij} P_X(\tau) [\prod_{m=1}^{k} (1 - Q_m(t-\tau))^{x'_m} \prod_{m=1}^{k} (1 - Q_m(t-\tau-1))^{s_m}$$

$$- \prod_{m=1}^{k} (1 - Q_m(t-\tau-1))^{x'_m} \prod_{m=1}^{k} (1 - Q_m(t-\tau-2))^{s_m}]. \tag{28}$$

Здесь $X' = (x'_1, \ldots, x'_k) = (x_1, \ldots, x_{i-1}, x_i - 1, x_{i+1}, \ldots, x_k)$ и s_m равно числу нетерминалов A_m в правой части правила r_{ij} $(m = 1, \ldots, k)$. Выражение в квадратных скобках в (28) аналогично выражению $R_X(t-\tau)$. Множителями $(1 - Q_m(t - \tau - 1))^{s_m}$ и $(1 - Q_m(t - \tau - 2))^{s_m}$ учитывается тот факт, что к первому нетерминалу A_i на ярусе τ применено правило r_{ij}, которому на ярусе $\tau + 1$ соответствует s_m вершин, помеченных нетерминалом A_m.

После несложных преобразований в (28) получаем

$$P\left(D^t_{X,1}(\tau)\right) = p_{ij} P_X(\tau) \frac{\prod\limits_{l=1}^{k} (1 - Q_l(t - \tau - 1))^{s_l}}{1 - Q_i(t - \tau)}$$

$$\times \left[\prod_{m=1}^{k} (1 - Q_m(t - \tau))^{x_m} - \prod_{m=1}^{k} (1 - Q_m(t - \tau - 1))^{x_m} \Delta \right],$$

где

$$\Delta = \frac{1 - Q_i(t - \tau)}{1 - Q_i(t - \tau - 1)} \prod_{m=1}^{k} \frac{(1 - Q_m(t - \tau - 2))^{s_m}}{(1 - Q_m(t - \tau - 1))^{s_m}}.$$

Из леммы 2 следует, что

$$\frac{1 - Q_l(n)}{1 - Q_l(n-1)} = 1 + \frac{P(D^n_l)}{1 - Q_l(n-1)} = 1 + O\left(\frac{1}{n^2}\right)$$

и

$$\left(\frac{1 - Q_l(n-1)}{1 - Q_l(n)}\right)^{s_m} = \left(1 - \frac{P(D^n_l)}{1 - Q_l(n)}\right)^{s_m} = 1 + O\left(\frac{1}{n^2}\right).$$

Поэтому $\Delta = 1 + O\left(\frac{1}{(t-\tau)^2}\right)$ и

$$P(D^t_{X,1)}(\tau)) = p_{ij} P_X(\tau) R_X(t - \tau)\Big(1 + O\Big(\tfrac{1}{t-\tau}\Big)\Big) + p_{ij} P_X(\tau) O\Big(\tfrac{1}{(t-\tau)^2}\Big).$$

Далее имеем

$$M_{ij}(t, \tau) = \frac{1}{P(D^t_1)} \sum_{X \neq 0} p_{ij} P_X(\tau) R_X(t - \tau) \left(1 + O\left(\frac{1}{t - \tau}\right)\right) x_i$$

$$+ \frac{1}{P(D^t_1)} \sum_{X \neq 0} p_{ij} P_X(\tau) x_i O\left(\frac{1}{(t - \tau)^2}\right).$$

Величина
$$\frac{1}{P(D_1^t)}\sum_{X\neq 0}P_X(\tau)R_X(t-\tau)x_i$$
есть $M_i(t,\tau)$ из теоремы 1, а величина $\sum\limits_{X\neq 0}P_X(\tau)x_i$ равна $a_{1i}^{(\tau)} = O(1)$ [2], где $a_{1i}^{(\tau)}$ — элемент матрицы A^τ и A — матрица первых моментов. Следовательно,
$$M_{ij}(t,\tau) = M_i(t,\tau)p_{ij}\left(1+O\left(\frac{1}{t-\tau}\right)\right)+O\left(\left(\frac{t}{t-\tau}\right)^2\right).$$
Применяя теорему 1 к $M_i(t,\tau)$, получаем
$$M_{ij}(t,\tau) = \frac{p_{ij}v_iB\tau(t-\tau)}{t}\left(1+O(\varepsilon)\right)+O\left(\left(\frac{t}{t-\tau}\right)^2\right)$$
$$= \frac{p_{ij}v_iB\tau(t-\tau)}{t}\left(1+O(\varepsilon)\right)$$
при $t\to\infty$ и $t\sqrt{\varepsilon}\leqslant\tau\leqslant t(1-\sqrt{\varepsilon})$.

Заметим, что при получении оценки для $M_{ij}(t,\tau)$ мы воспользовались (7), а также тем, что суммируется конечное число бесконечно малых величин, зависящих от t, τ и $t-\tau$. Поэтому
$$M_{ij}(t,\tau) = \frac{p_{ij}v_iB\tau(t-\tau)}{t}\left(1+\chi_{ij}(t,\tau,\varepsilon)\right),$$
где $|\chi_{ij}(t,\tau,\varepsilon)|\leqslant c_0\varepsilon$ при некоторой константе c_0, не зависящей от t и τ. Теорема доказана.

Пусть $S_{ij}(t) = q_{ij}(t,1)+q_{ij}(t,2)+\ldots+q_{ij}(t,t)$, где $q_{ij}(t,\tau)$ — число применений правила r_{ij} на ярусе τ в дереве вывода из D_1^t.

Теорема 3. *При $t\to\infty$ выполняется асимптотическое равенство*
$$M\left(S_{ij}(t)\right)\sim\frac{p_{ij}v_iBt^2}{6}.$$

Доказательство. Возьмем ε из интервала $(0,1/4)$. Положим $\tau_1 = \lfloor t\sqrt{\varepsilon}\rfloor$ и $\tau_2 = \lfloor t\left(1-\sqrt{\varepsilon}\right)\rfloor$. Разобьем $S_{ij}(t)$ на три части:
$$S_{ij}(t) = S_{ij}^{(1)}(t)+S_{ij}^{(2)}(t)+S_{ij}^{(3)}(t),$$

где $S_{ij}^{(1)}(t) = \sum_{\tau=1}^{\tau_1} q_{ij}(t,\tau)$, $S_{ij}^{(2)}(t) = \sum_{\tau=\tau_1+1}^{\tau_2} q_{ij}(t,\tau)$ и

$S_{ij}^{(3)}(t) = \sum_{\tau=\tau_2+1}^{t} q_{ij}(t,\tau)$. Оценим математические ожидания $M\left(S_{ij}^{(1)}(t)\right)$, $M\left(S_{ij}^{(2)}(t)\right)$ и $M\left(S_{ij}^{(3)}(t)\right)$.

Величину $M\left(S_{ij}^{(1)}(t)\right)$ можно представить в виде:

$$M\left(S_{ij}^{(1)}(t)\right) = M_{ij}(t,1) + M_{ij}(t,2) + \ldots + M_{ij}(t,\tau_1).$$

Для оценки $M_{ij}(t,\tau)$ при $\tau \leqslant \tau_1$ учтем, что $q_{ij}(t,\tau)$ не превосходит числа вершин на ярусе τ, помеченных нетерминалом A_i. Поэтому

$$M_{ij}(t,\tau) \leqslant M_i(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} P_X(\tau) R_X(t-\tau) x_i.$$

Применяя леммы 4 и 5, получаем

$$R_X(t-\tau) \leqslant c_1(t-\tau) \sum_{m=1}^{k} P(D_m^{t-\tau}) \leqslant \frac{c_2}{t-\tau},$$

где c_1 и c_2 — некоторые константы. Следовательно,

$$M_{ij}(t,\tau) \leqslant \frac{c_2}{P(D_1^t)(t-\tau)} \sum_{X} P_X(\tau) x_i \leqslant c\frac{t^2}{t-\tau},$$

так как $\sum\limits_{X} P_X(\tau) x_i = a_{1i}^{(\tau)} = O(1)$ (здесь c — некоторая константа).

Число слагаемых в $S_{ij}^{(1)}(t)$ равно $\tau_1 \leqslant t\sqrt{\varepsilon}$, а $t-\tau \geqslant t/2$. Поэтому

$$M\left(S_{ij}^{(1)}(t)\right) \leqslant 2c\sqrt{\varepsilon}t^2 = O(t^2\sqrt{\varepsilon}).$$

Рассмотрим τ, удовлетворяющее условию $\tau_2 + 1 \leqslant \tau \leqslant t$. Для любого такого τ имеем

$$M_{ij}(t,\tau) \leqslant M_i(t,\tau) = \frac{1}{P(D_1^t)} \sum_{X \neq 0} P_X(\tau) x_i R_X(t-\tau).$$

Для оценки $R_X(t-\tau)$ применим (6):

$$R_X(t-\tau) \leqslant \sum_{l=1}^{k} x_l P(D_l^{t-\tau}) \left(1 + \varphi_l(t-\tau)\right) \prod_{m=1}^{k} \left(1 - Q_m(t-\tau)\right)^{x_m}$$

$$\leqslant 2\sum_{l=1}^{k} x_l P(D_l^{t-\tau}) \prod_{m=1}^{k} (1-Q_m(t-\tau))^{x_m}.$$

Используя лемму 5 и учитывая неравенство $1-Q_m(t-\tau)<1$, получаем

$$M_{ij}(t,\tau) \leqslant \frac{2}{P(D_1^t)} \sum_{X\neq 0} P_X(\tau) \sum_{l=1}^{k} \frac{4P(D_l^{t-\tau})}{Q_l(t-\tau)Q_i(t-\tau)}.$$

Рассмотрим функцию $f_l(n) = \dfrac{P(D_l^n)}{Q_l(n)Q_i(n)}$. Из лемм 1 и 2 следует, что $\lim\limits_{n\to\infty} f_l(n) = \dfrac{B}{2u_i}$. Кроме того, $f_l(n)$ определена при любом натуральном n, так как $Q_l(t-\tau)>0$ при $l=1,\ldots,k$ в силу бесконечности языка. Значит, $f_l(n)$ ограничена некоторой константой $c_l>0$. Поэтому

$$M_{ij}(t,\tau) \leqslant \frac{8}{P(D_1^t)} \sum_{l=1}^{k} c_l \sum_{X\neq 0} P_X(\tau) = \frac{8Q_1(\tau)}{P(D_1^t)} \sum_{l=1}^{k} c_l.$$

Очевидно, что $Q_1(\tau) \leqslant Q_1(\tau_2)$ при $\tau>\tau_2$ и

$$M_{ij}(t,\tau) \leqslant \frac{8Q_1(\tau_2)}{P(D_1^t)} \sum_{l=1}^{k} c_l \leqslant \frac{ct^2}{\tau_2} \leqslant \frac{ct}{1-\sqrt{\varepsilon}}$$

(здесь c — некоторая константа).

В силу выбора ε из интервала $(0,1/4)$ значение $1-\sqrt{\varepsilon}$ больше $1/2$. Поэтому $M_{ij}(t,\tau) \leqslant 2ct$. Так как число слагаемых в $S_{ij}^{(3)}(t)$ не превосходит $\sqrt{\varepsilon}t$, то

$$M\left(S_{ij}^{(3)}(t)\right) \leqslant 2\sqrt{\varepsilon}ct^2 = O\left(t^2\sqrt{\varepsilon}\right).$$

Наконец, найдем значение $M\left(S_{ij}^{(2)}(t)\right)$, применив к $M_{ij}(t,\tau)$ теорему 2, так как в области $\tau_1<\tau\leqslant\tau_2$ выполняются ограничения на τ, необходимые для ее применения. Можно записать

$$M\left(S_{ij}^{(2)}(t)\right) = \sum_{\tau=\tau_1+1}^{\tau_2} M_{ij}(t,\tau) = \frac{p_{ij}v_iB}{t} \sum_{\tau=\tau_1+1}^{\tau_2} \{\tau(t-\tau)(1+\chi_{ij}(t,\tau,\varepsilon))\}.$$

Очевидно, что

$$\sum_{\tau=\tau_1+1}^{\tau_2} \tau = \frac{t^2}{2}\left(1+O\left(\sqrt{\varepsilon}\right)\right).$$

Для $\sum\limits_{\tau=\tau_1+1}^{\tau_2} \tau^2$ справедливы неравенства

$$\int_{\tau_1+1}^{\tau_2+1} (\tau-1)^2 d\tau \leqslant \sum_{\tau=\tau_1+1}^{\tau_2} \tau^2 \leqslant \int_{\tau_1+1}^{\tau_2+1} \tau^2 d\tau.$$

Поэтому вычисляя интегралы, получаем

$$\sum_{\tau=\tau_1+1}^{\tau_2} \tau^2 = \frac{t^3}{3}\left(1 + O\left(\sqrt{\varepsilon}\right)\right).$$

Следовательно,

$$\sum_{\tau=\tau_1+1}^{\tau_2} \tau(t-\tau) = \frac{t^3}{6}\left(1 + O\left(\sqrt{\varepsilon}\right)\right).$$

С учетом оценки для $\chi_{ij}(t,\tau,\varepsilon)$ из теоремы 2, находим

$$M\left(S_{ij}^{(2)}(t)\right) = \frac{p_{ij}v_iBt^2}{6}\left(1 + O\left(\sqrt{\varepsilon}\right) + O\left(\varepsilon\right)\right) = \frac{p_{ij}v_iBt^2}{6}\left(1 + O\left(\sqrt{\varepsilon}\right)\right).$$

Суммируя оценки для $M\left(S_{ij}^{(1)}(t)\right)$, $M\left(S_{ij}^{(2)}(t)\right)$ и $M\left(S_{ij}^{(3)}(t)\right)$, получаем

$$M\left(S_{ij}(t)\right) = \frac{p_{ij}v_iBt^2}{6}\left(1 + O(\sqrt{\varepsilon})\right).$$

Так как это равенство выполняется для любого сколь угодно малого $\varepsilon > 0$, справедлива следующая асимптотика при $t \to \infty$:

$$M\left(S_{ij}(t)\right) \sim \frac{p_{ij}v_iBt^2}{6}.$$

Теорема 3 доказана.

4. Нижняя оценка стоимости кодирования

В этом разделе будем рассматривать грамматики с однозначным выводом, т. е. грамматики, в каждой из которых любое слово порождаемого языка имеет единственное дерево вывода.

Пусть L — стохастический КС-язык. Через L^t обозначим множество таких слов из L, что дерево вывода каждого слова имеет высоту t.

Для $\alpha \in L^t$ через $p_t(\alpha)$ обозначим условную вероятность появления слова α, т. е. $p_t(\alpha) = \dfrac{p(\alpha)}{P(L^t)}$. В силу однозначности вывода $P(L^t) = P(D^t)$.

Кодированием языка L будем называть инъективное отображение $f : L \to \{0,1\}^+$. *Стоимостью кодирования* f назовем величину

$$C(L,f) = \lim_{t\to\infty} \frac{\sum\limits_{\alpha\in L^t} p_t(\alpha)|f(\alpha)|}{\sum\limits_{\alpha\in L^t} p_t(\alpha)|\alpha|} \tag{29}$$

(здесь $|x|$ -длина последовательности x).

Величина $C(L,f)$ равна среднему числу двоичных разрядов, используемых при кодировании одного символа слова.

Через $F(L)$ обозначим множество таких инъективных отображений f из L в $\{0,1\}^+$, что существует $C(L,f)$.

Стоимостью *оптимального* кодирования языка L назовем величину

$$C_0(L) = \inf_{f\in F(L)} C(L,f). \tag{30}$$

Под *энтропией* множества слов L^t будем понимать величину

$$H(L^t) = -\sum_{\alpha\in L^t} p_t(\alpha)\log p_t(\alpha).$$

(Здесь и далее логарифм берется по основанию 2.)

Теорема 4. *Пусть L — язык, порожденный стохастической* КС-*грамматикой с однозначным выводом, матрица первых моментов которой неразложима, непериодична и ее перронов корень равен* 1. *Тогда при* $t\to\infty$

$$H(L^t) \sim \frac{Bt^2}{6}\sum_{i=1}^{k} v_i H(R_i), \tag{31}$$

где $V = (v_1,\dots,v_k)$ *— левый собственный вектор матрицы первых моментов, соответствующий перронову корню, и* $H(R_i)$ *— энтропия множества правил* R_i, $H(R_i) = -\sum\limits_{j=1}^{n_i} p_{ij}\log p_{ij}$.

Доказательство. Обозначим через $q_{ij}(\alpha)$ число применений правила r_{ij} в выводе слова α из L^t. Тогда

$$p(\alpha) = \prod_{i=1}^{k}\prod_{j=1}^{n_j} p_{ij}^{q_{ij}}(\alpha) \quad \log p(\alpha) = \sum_{i=1}^{k}\sum_{j=1}^{n_j} q_{ij}(\alpha)\log p_{ij}.$$

Очевидно, что для логарифма условной вероятности $p_t(\alpha)$ справедлива формула $\log p_t(\alpha) = \log p(\alpha) - \log P(L^t)$. Используя полученные формулы, проведем преобразования $H(L^t)$:

$$H(L^t) = -\sum_{\alpha\in L^t} p_t(\alpha)\log p_t(\alpha) = \frac{1}{P(L^t)}\left(-\sum_{\alpha\in L^t} p(\alpha)\left(\log p(\alpha) - \log P(L^t)\right)\right)$$

$$= \frac{1}{P(L^t)}\left(-\sum_{i=1}^{k}\sum_{j=1}^{n_j}\log p_{ij}\sum_{\alpha\in L^t} p(\alpha)q_{ij}(\alpha)\right) + \log P(L^t).$$

Так как $\log P(L^t) = O(\log t)$ по лемме 2 и

$$\sum_{\alpha\in L^t} p(\alpha)q_{ij}(\alpha) = P(L^t)M(S_{ij}(t)) = P(L^t)\frac{p_{ij}v_iBt^2}{6}(1+o(1))$$

по теореме 3, то

$$H(L^t) = \frac{Bt^2}{6}(1+o(1))\left(-\sum_{i=1}^{k} v_i\sum_{j=1}^{n_j} p_{ij}\log p_{ij}\right) + O(\log t)$$

$$= \frac{Bt^2}{6}(1+o(1))\left(\sum_{i=1}^{k} v_iH(R_i)\right) + O(\log t) \sim \frac{Bt^2}{6}\sum_{i=1}^{k} v_iH(R_i).$$

Теорема 4 доказана.

Теорема 5. *Пусть L — язык, порожденный стохастической* КС-*грамматикой с однозначным выводом, матрица первых моментов которой неразложима, непериодична и перронов корень равен* 1. *Тогда*

$$C_0(L) = -\frac{1}{h}\sum_{i=1}^{k} v_i\sum_{j=1}^{n_i} p_{ij}\log p_{ij},$$

где $C_0(L)$ определено в (30), $h = \sum_{i=1}^{k} v_i\sum_{j=1}^{n_i} p_{ij}l_{ij}$, p_{ij} — *вероятность правила r_{ij}, $V = (v_1,\ldots,v_k)$ — левый собственный вектор для перронова корня r и l_{ij} — число терминальных символов в правой части правила r_{ij}.*

Доказательство. Рассмотрим способ кодирования слов из L^t, состоящий в упорядочении слов в порядке невозрастания вероятностей их появления и кодировании их по порядку сначала двоичными словами длины 1, затем двоичными словами длины 2, и т. д. Такое кодирование обозначим через f^*. Очевидно, что при таком f^* сумма $\sum_{\alpha \in L^t} p_t(\alpha)|f^*(\alpha)|$ минимальна среди всех возможных кодирований множества слов из L^t. Поэтому для любого кодирования f слов языка L, включающем множество L^t, выполняется неравенство

$$\sum_{\alpha \in L^t} p_t(\alpha)|f(\alpha)| \geqslant \sum_{\alpha \in L^t} p_t(\alpha)|f^*(\alpha)|.$$

Пусть $M_t(f^*) := \sum_{\alpha \in L^t} p_t(\alpha)|f^*(\alpha)|$.

В [5] доказана нижняя оценка стоимости кодирования слов из конечного множества с заданным на нем распределением вероятностей, где под стоимостью кодирования понимается математическое ожидание длины закодированного слова. Применяя эту оценку к множеству слов из L^t, получаем

$$M_t(f^*) \geqslant H(L^t) - \log\log N - C, \tag{32}$$

где N — число слов в множестве L^t и C — некоторая константа.

Оценим сверху величину N. Для грамматики с однозначным выводом число слов в L^t равно числу различных деревьев вывода высоты t. Из каждой вершины дерева вывода выходит не более c_1 дуг, где c_1 равно наибольшей длине слова в правой части правила грамматики. Поэтому число вершин на ярусе τ не превосходит c_1^τ. Общее число вершин в дереве вывода высоты t, помеченных нетерминалами, не превосходит

$$\sum_{\tau=0}^{t-1} c_1^\tau \leqslant c_1^t.$$

Дерево вывода высоты t можно закодировать последовательностью номеров правил грамматики в левом выводе соответствующего слова из L^t. Очевидно, длина этой последовательности не меньше t и не превосходит c_1^t.

На каждом месте в левом выводе может стоять не более c_2 различных номеров правил грамматики, где c_2 — общее число правил в грамматике. Поэтому число N деревьев вывода высоты t не превосходит величины

$$\sum_{n=t}^{c_1^t} c_2^n \leqslant c_2^{c_1^t+1}.$$

Значит, $\log\log N \leqslant \log\log c_2^{c_1^t+1} \leqslant t\log c_1 + \log\log c_2 + 1 = O(t)$.

Учитывая найденную оценку для N и применяя теорему 4, неравенство (32) можно переписать в виде:

$$M_t(f^*) \geqslant H(L^t) + O(t) \geqslant H(L^t)\left(1 + O\left(\frac{1}{t}\right)\right).$$

Теперь вычислим $M(|\alpha|)$, когда $\alpha \in L^t$. Для этого длину слова α представим в виде:

$$|\alpha| = \sum_{i=1}^{k}\sum_{j=1}^{n_i} q_{ij}(\alpha) l_{ij},$$

где l_{ij} — число терминальных символов в правой части правила r_{ij}. Тогда при $\alpha \in L^t$ имеем

$$M(|\alpha|) = \sum_{\alpha\in L^t} p_t(\alpha)\sum_{i=1}^{k}\sum_{j=1}^{n_i} q_{ij}(\alpha) l_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} l_{ij} M(S_{ij}(t)).$$

Применив теорему 3 к $M(S_{ij}(t))$, получим

$$M(|\alpha|) = \frac{Bt^2}{6}(1 + o(1))\sum_{i=1}^{k} v_i \sum_{j=1}^{n_i} p_{ij} l_{ij}.$$

Пусть $h := \sum\limits_{i=1}^{k} v_i \sum\limits_{j=1}^{n_i} p_{ij} l_{ij}$. Тогда

$$C(L,f) = \lim_{t\to\infty}\frac{\sum\limits_{\alpha\in L^t} p_t(\alpha)|f(\alpha)|}{\sum\limits_{\alpha\in L^t} p_t(\alpha)|\alpha|} \geqslant \lim_{t\to\infty}\frac{M_t(f^*)}{M(|\alpha|)} = \frac{6M_t(f^*)}{Bt^2 h}.$$

Наконец, воспользовавшись (32), а затем (31), получаем

$$C(L,f) \geqslant C_0(L) \geqslant \frac{1}{h}\sum_{i=1}^{n} v_i H(R_i).$$

Полученная нижняя оценка неулучшаема, поскольку из доказательства теоремы Шеннона для канала без шума [7] следует неравенство

$$M_t(f^*) \leqslant H(L^t) + 1.$$

Следовательно,

$$C_0(L) \leqslant C(L, f^*) \leqslant \frac{1}{h} \sum_{i=1}^{n} H(R_i).$$

Теорема 5 доказана.

Литература

1. **Ахо А., Ульман Дж.** Теория синтаксического анализа, перевода и компиляции. Том 1. М.: Мир, 1978.
2. **Гантмахер Ф. Р.** Теория матриц. М.: Наука, 1967.
3. **Жильцова Л. П.** Закономерности применения правил грамматики в выводах слов стохастического контекстно-свободного языка// Математические вопросы кибернетики. Вып.9. М.: Наука, 2000. С. 101–126.
4. **Жильцова Л. П.** О нижней оценке стоимости кодирования и асимптотически оптимальном кодировании стохастического контекстно-свободного языка// Дискрет. анализ и исслед. операций. Сер. 1. 2001. Т. 8, № 3. С. 26–45.
5. **Кричевский Р. Е.** Сжатие и поиск информации. М.: Радио и связь, 1989.
6. **Севастьянов В. А.** Ветвящиеся процессы. М.: Наука, 1971.
7. **Шеннон К.** Математическая теория связи. М.: ИЛ, 1963.
8. **Ширяев А. Н.** Вероятность. М.: Наука, 1980.

Адрес автора:

Нижегородский государственный
педагогический университет,
ул. Ульянова, 1,
603005 Нижний Новгород,
Россия

Статья поступила
29 апреля 2003 г.