

# Курсовая работа

по теме:

«Энтропия множества деревьев вывода в  
разложимой стохастической  
КС-грамматике, имеющей вид "цепочки"»

**Студент:**

Матрынов Игорь Михайлович

**Научный руководитель:**

Жильцова Лариса Павловна

# Стохастическая КС-грамматика

$$G = \langle V_N, V_T, s, R \rangle$$

$V_N$  и  $V_T$  — конечные алфавиты терминальных и нетерминальных символов

$s \in V_N$  — аксиома

$R = \cup_{i=1}^k R_i$ , где  $R_i$  — множество правил вида:

$$r_{ij} : A_i \xrightarrow{p_{ij}} \beta_{ij} \quad (j = 1, 2, \dots, n_i)$$

$\beta_{ij} \in (V_N \cup V_T)^*$ ,  $p_{ij}$  — вероятность применения правила

$$0 < p_{ij} \leq 1 \quad \sum_{j=1}^{n_i} p_{ij} = 1$$

# Пример вывода слова

Язык арифметических выражений (+, -, \*, /, a, b)

$$V_N = \{E, S, M\}, V_T = \{a, b\}, s = E$$

$$a \cdot (b + a) - b$$

Множество правил вывода

$$r_{11} : E \rightarrow E + S$$

$$r_{12} : E \rightarrow E - S$$

$$r_{13} : E \rightarrow S$$

$$r_{21} : S \rightarrow S \cdot M$$

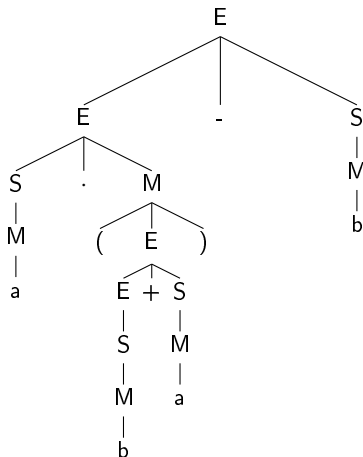
$$r_{22} : S \rightarrow S / M$$

$$r_{23} : S \rightarrow M$$

$$r_{31} : M \rightarrow (E)$$

$$r_{32} : M \rightarrow a$$

$$r_{33} : M \rightarrow b$$



# Отношение следования на множестве нетерминалов

$A_i \rightarrow A_j$ , если существует правило  $A_i \xrightarrow{p_{ij}} \alpha A_j \beta$ , где  $\alpha, \beta \in (V_N \cup V_T)^*$ .

$(\rightarrow_*)$  — рефлексивное транзитивное замыкание  $(\rightarrow)$ .

$A_i \leftrightarrow_* A_j$ , если  $A_i \rightarrow_* A_j$  и  $A_j \rightarrow_* A_i$ .

Разобьём нетерминалы на классы по отношению  $(\leftrightarrow_*)$ . Тогда:

$$\forall A_i, A_j \in K \quad A_i \rightarrow_* A_j$$

Не рассматриваем особые классы:  $K = \{A\}$ ,  $A \not\rightarrow A$ .

$K_1 \prec K_2$ , если  $\exists A_1 \in K_1, A_2 \in K_2 : A_1 \rightarrow A_2$ .

$(\prec_*)$  — рефлексивное транзитивное замыкание  $(\prec)$ .

# Производящие функции

$$F_i(s_1, s_2, \dots, s_k) = \sum_{j=1}^{n_i} p_{ij} s_1^{l_1} s_2^{l_2} \dots s_k^{l_k},$$

где  $l_s$  — число нетерминалов  $A_s$  в правой части правила  $A_i \xrightarrow{p_{ij}} \beta_{ij}$ .

Производящие функции с параметром  $t$ :

$$F_i(t, \bar{s}) = \begin{cases} F_i(\bar{s}), & \text{при } t = 1 \\ F_i(\bar{F}(t-1, \bar{s})), & \text{при } t = 2, 3, \dots \end{cases}$$

## Первые моменты

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

$a_j^i$  — математическое ожидание числа нетерминалов  $A_j$  при однократном применении случайного правила к  $A_i$ .

Матрица  $A = (a_j^i)$  — матрица первых моментов.

$r$  — максимальное по модулю собственное число матрица  $A$  (перронов корень). Будем рассматривать критический случай  $r = 1$ .

## Пример: язык арифметических выражений (без скобок)

$$V_N = \{E, S, M, X\} \quad V_T = \{a, b, c\} \quad s = E$$

$R$  содержит правила:

$$r_{11} : E \xrightarrow{1/2} E + E$$

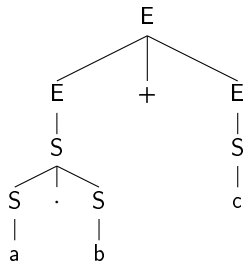
$$r_{12} : E \xrightarrow{1/2} S$$

$$r_{21} : S \xrightarrow{1/4} S \cdot S$$

$$r_{22} : S \xrightarrow{1/4} a$$

$$r_{23} : S \xrightarrow{1/4} b$$

$$r_{24} : S \xrightarrow{1/4} c$$

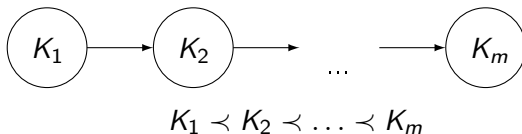


Матрица первых моментов:

$$\begin{pmatrix} 1 & 1/2 \\ 0 & 1/2 \end{pmatrix}$$

$$\begin{aligned} p(a \cdot b + c) &= p_{11} \cdot p_{12} \cdot p_{21} \cdot p_{22} \cdot p_{23} \cdot p_{12} \cdot p_{24} = \\ &= \left(\frac{1}{2}\right)^3 \cdot \left(\frac{1}{4}\right)^4 = \frac{1}{2048} \end{aligned}$$

## Грамматика вида «цепочки»



Матрица первых моментов такой грамматики имеет вид:

$$A = \begin{pmatrix} A_{11} & A_{12} & 0 & \dots & 0 & 0 \\ 0 & A_{22} & A_{23} & \dots & 0 & 0 \\ 0 & 0 & A_{33} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & A_{m-1,m-1} & A_{m-1,m} \\ 0 & 0 & 0 & \dots & 0 & A_{mm} \end{pmatrix}$$

Блоки  $A_{ii} (i = 1, 2, \dots, m)$  неразложимы, имеют перроновы корни  $r_i$ .

Перронов корень всей матрицы  $r = \max\{r_i\}$ .



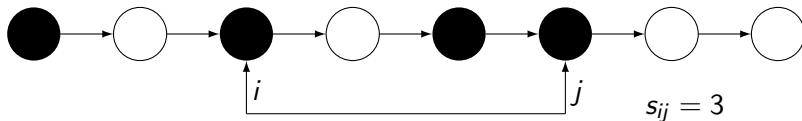
# Асимптотика матрицы первых моментов

Пусть  $A_{ij}^{(t)}$  — блок матрицы  $A^t$  на позиции блока  $A_{ij}$  матрицы  $A$ .  
При  $t \rightarrow \infty$

$$A_{ij}^{(t)} = U^{(i)} V^{(j)} \cdot t^{s_{ij}-1} \cdot (1 + o(1))$$

Если класс  $K_i$  ( $K_j$ ) критический, то  $U^{(i)}$  ( $V^{(i)}$ ) — правый (левый) собственный вектор блока  $A_{ii}$  ( $A_{jj}$ ).

*Л. П. Жильцова*



## Вероятности продолжения

$Q_i(t)$  — вероятность деревьев вывода высоты  $\geq t$  с корнем  $A_i$  (вероятности продолжения).

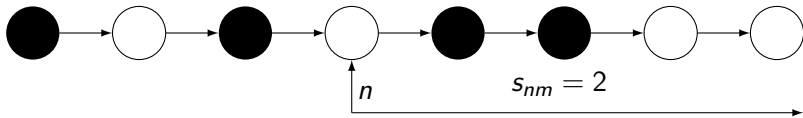
$P_i(t)$  — вероятность деревьев высоты  $t$ .

При  $t \rightarrow \infty$

$$Q_i(t) = c_n \cdot U_i \cdot t^{-\left(\frac{1}{2}\right)^{s_{nm}-1}} \cdot (1 + o(1))$$

$$P_i(t) = \tilde{c}_n \cdot U_i \cdot t^{-1-\left(\frac{1}{2}\right)^{s_{nm}-1}} \cdot (1 + o(1))',$$

где  $A_i \in K_n$ ,  $U_i, c_n, \tilde{c}_n > 0$ .



# Математические ожидания числа применений правил

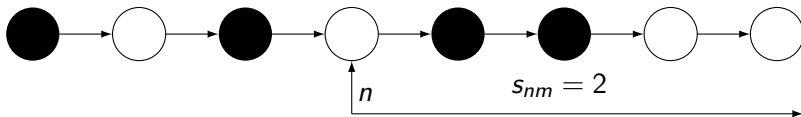
$M_{ij}(t)$  — математическое ожидание числа применений правила  $r_{ij}$  в деревьях высоты  $t$

При  $t \rightarrow \infty$

$$M_{ij}(t) = d_i \cdot p_{ij} \cdot t^{\left(\frac{1}{2}\right)^{s_{nm}-1-\xi}} \cdot (1 + o(1)),$$

где  $A_i \in K_n$ ,  $d_i > 0$  и

$$\xi = \begin{cases} 1, & K_n \text{ — критический} \\ 0, & K_n \text{ — докритический} \end{cases}$$



# Энтропия деревьев вывода фиксированной высоты

При  $t \rightarrow \infty$

$$H(D^t) = t^2 \cdot \sum_{i > \sigma_{n-1}} d_i \sum_{j=1}^{n_i} p_{ij} \log p_{ij} \cdot (1 + o(1)),$$

где  $\sigma_{n-1}$  — число нетерминалов в классах, предшествующих последнему критическому

В докритическом случае:

$$H(D^t) = t \cdot \left( \log r - \sum_{i=1}^k \sum_{j=1}^{n_i} w_{ij} \log p_{ij} \right) \cdot (1 + o(1))$$

# Использованная литература

- ▶ **Севастьянов Б. А.** Ветвящиеся процессы. — М.: Наука, 1971 — 436 с.
- ▶ **Фу К.** Структурные методы в распознавании образов. М.: Мир, 1977
- ▶ **Гантмахер Ф. Р.** Теория матриц. — 5-е изд., — М.: ФИЗМАТЛИТ, 2010
- ▶ **Ахо А., Ульман Дж.** Теория синтаксического анализа, перевода и компиляции. Том 1. М.: Мир, 1978
- ▶ **Жильцова Л. П.** О матрице первых моментов разложимой стохастической КС-грамматики. УЧЁНЫЕ ЗАПИСКИ КАЗАНСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА, Том 151, кн. 2, 2009
- ▶ **Жильцова Л. П.** Закономерности применения правил грамматики в выводах слов стохастического контекстно-свободного языка // Математические вопросы кибернетики. Выр. 9. М.: Наука, 2000. С. 100-126.
- ▶ **Борисов А. Е.** Закономерности в словах стохастических контекстно-свободных языков, порождённых грамматиками с двумя классами нетерминальных символов. Вопросы экономного кодирования.

# Спасибо за внимание!