Вид матрицы первых моментов для грамматик в виде «цепочки»

Игорь Мартынов

5 марта 2011 г.

1 Основные определения

Стохастической КС-грамматикой называется система $G =< V_T, V_N, R, s >$, где V_T и V_N — алфавиты терминальных и нетерминальных символов соответственно, s — аксиома грамматики, R — множество правил вывода, представимое в виде $R = \cup_{i=1}^{k} R_i$, где $k = |V_N|$, и R_i — множество правил вида

$$r_{ij} : A_i \xrightarrow{p_{ij}} \beta_{ij} \quad (A_i \in V_N, \beta_{ij} \in (V_N \cup V_T)^*)\,, \tag{1}$$

и p_{ij} — вероятность применения правила r_{ij}, причём при фиксированном i вероятности r_{ij} задают вероятностное распределение на множестве R_i:

$$0 < p_{ij} \leq 1 \quad \text{и} \quad \sum_{j=1}^{n_i} p_{ij} = 1, \qquad i = 1, 2, \ldots, k, \tag{2}$$

где $n_i = |R_i|$.

Слово β называется *непосредственно выводимым* из α (обозначается $\alpha \Rightarrow \beta$), если существуют $\alpha_1, \alpha_2 \in (V_N \cup V_N)^*$, для которых $\alpha = \alpha_1 A_i \alpha_2$, $\beta = \alpha_1 \beta_{ij} \alpha_2$ и в R имеется правило $A_i \xrightarrow{p_{ij}} \beta_{ij}$.

Через $\Rightarrow_*$ обозначим рефлексивное транзитивное замыкание $\Rightarrow$. Если $\alpha \Rightarrow_* \beta$, говорят, что β *выводимо* из α. Язык, *порождаемый* грамматикой G определяется как $L_G = \{\alpha : s \Rightarrow_* \alpha, \alpha \in V_T^*\}$.

Последовательность правил грамматики $\omega(\alpha) = (r_1, r_2, \ldots, r_\gamma)$, последовательное применение которых к s даёт слово α, называется *выводом* этого слова. Если на каждом шаге правило применяется к самому левому нетерминалу в слове, вывод назвается *левым*.

Вероятность вывода определяется как $p(\omega(\alpha)) = p(r_1) \cdot p(r_2) \cdot \ldots \cdot p(r_\gamma)$, где $p(r_i)$ — вероятность соответствующего правила. Вероятность слова определяется как сумма вероятностей всех его левых выводов.

Грамматика G называется *согласованной*, если

$$\sum_{\alpha \in L_G} p(\alpha) = 1. \tag{3}$$

Согласованная грамматика G задаёт распределение вероятностей P на L_G, и определяет *стохастический КС-язык* $\mathfrak{L} = (L, P)$. В дальнейшем всюду предполагается, что грамматика согласованна.

По выводу слова может быть построено *дерево вывода.* В узел дерева помещается аксиома s, далее на каждом ярусе дерева ко всем нетерминалам этого яруса применяется правило, соответствующее выводу. Символы этого слова записываются слева направо в дереве, присоединяясь к исходному нетерминалу как к родителю.

Для исследования вероятностных характеристик стохастической грамматики применяются производящие функции

$$F_i(s_1, s_2, \dots, s_k) = \sum_{\substack{j=1 \\ r_{ij} \in R}}^{n_i} p_{ij} s_1^{l_1} s_2^{l_2} \dots s_k^{l_k}, \tag{4}$$

где $l_m = l_m(i, j)$ — число вхождений нетерминала A_m в β_{ij}.

Величины

$$a_j^i = \left. \frac{\partial F_i(s_1, s_2, \dots, s_k)}{\partial s_j} \right|_{s_1 = s_2 = \dots = s_k = 1} \tag{5}$$

называются *первыми моментами* грамматики G. Матрица $A = (a_j^i)$, составленная из них, называется *матрицей первых моментов* грамматики G.

Матрица A, по построению, неотрицательна. По теореме Фробениуса, существует максимальный по модулю вещественный неотрицательный собственный корень r. Известно, что критерием согласованности стохастической КС-грамматики при отсутствии бесполезных нетерминалов является условие $r \leq 1$.

Говорят, что нетерминал A_j *непосредственно следует* за нетерминалом A_i (обозначается $A_i \to A_j$), если в R имеется правило $A_i \overset{(}{\to} p_{ij})\alpha_1 A_j \alpha_2$, где $\alpha_1, \alpha_2 \in (V_N \cup V_T)^*$. Транзитивное замыкание отношения $\to$ обозначается $\to_*$. Если $A_i \to_* A_j$, говорят, что A_j *выводится* из A_i.

Введём также отношение $\leftrightarrow_*$. Будем считать, что $A_i \leftrightarrow_* A_j$, если одновременно $A_i \to_* A_j$ и $A_j \to_* A_i$, либо если $A_i = A_j$. Очевидно, отношение $\leftrightarrow_*$ есть отношение эквивалентности, и потому разбивает множество нетерминалов на классы $V_N = K_1 \cup K_2 \cup \ldots \cup K_m : K_i \cap K_j = \varnothing (i \neq j)$. Класс, содержащий ровно один нетерминал, будем называть *особым.* Множество классов $\{K_1, K_2, \dots, K_m\}$ обозначим $\mathfrak{K}$.

Если все нетерминалы грамматики образуют один класс, она называется *неразложимой.* В противном случае она называется *разложимой.* Очевидно, разложимой грамматике соответствует разложимая матрица первых моментов.

Говорят, что класс K_j *непосредственно следует* за классом K_i (обозначается $K_i \prec K_j$), если существуют $A_1 \in K_i$ и $A_2 \in K_j$ такие, что $A_1 \to A_2$. Рефлексивное транзитивное замыкание $\prec$ обозначим $\prec_*$, и назовём отношением *следования.*

Будем говорить, что грамматика имеет вид *«цепочки»*, если она разложима, и граф, построенный на множестве $\mathfrak{K}$ по отношению $\prec$, имеет вид P_m. Пронумеруем классы грамматики таким образом, что $K_i \prec K_{i+1}, i = 1, 2, \dots, m-1$. Пронумеруем нетерминалы так, что для любых $A_i \in K_p$ и $A_j \in K_q$ $i < j \Leftrightarrow p < q$. После этого матрица первых моментов грамматики приобретает вид:

$$A = \begin{pmatrix} A_{11} & A_{12} & 0 & \cdots & 0 & 0 \\ 0 & A_{22} & A_{23} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_{m-1,m-1} & A_{m-1,m} \\ 0 & 0 & 0 & \cdots & 0 & A_{m,m} \end{pmatrix} \tag{6}$$

Блоки $A_{i,i}(i = 1, 2, \ldots, m)$ являются неразложимыми неотрицательными матрицами. Не уменьшая общности, будем считать их положительными и непериодичными. Этого можно добиться с помощью метода укрупнения правил грамматики. Пусть r_i — перронов корень матрицы $A_{i,i}$. По построению матрицы A, $r = \max_i\{r_i\}$ и $r > 0$.

2 Свойства матрицы первых моментов

Обозначим $J = \{i : r_i = r\} = \{i_1 < i_2 < \ldots < i_q\}$ Разобьём множество классов $\mathfrak{K}$ на группы классов $\mathfrak{M_1}, \mathfrak{M_2}, \ldots, \mathfrak{M_w}$. При этом $\mathfrak{M_1} = \{K_1, K_2, \ldots, K_{i_1}\}$, и $\mathfrak{M_l} = \{K_{i_{l-1}+1}, \ldots, K_{i_l}\}$, где $l > 1$. При таком разбиении в каждой группе $\mathfrak{M_j}$ содержится ровно один класс с номером из J.

Тогда матрицу первых моментов можно представить в виде

$$A = \begin{pmatrix} B_{11} & B_{12} & \cdots & 0 & 0 \\ 0 & B_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & B_{w-1,w-1} & B_{w-1,w} \\ 0 & 0 & \cdots & 0 & B_{w,w} \end{pmatrix}, \tag{7}$$

где B_{ij} — блок, находящийся на пересечении строк, соответствующих нетерминалам классов группы $\mathfrak{M_i}$, и столбцов, соответствующим нетерминалам классов группы $\mathfrak{M_j}$. Очевидно, каждой из матриц $B_{i,i}$ соответствует перронов корень равный r.

Рассмотрим матрицу

$$A^t = \begin{pmatrix} B_{11}^t & B_{12}^{(t)} & \cdots & B_{1,w-1}^{(t)} & B_{1,w}^{(t)} \\ 0 & B_{22}^t & \cdots & B_{2,w-1}^{(t)} & B_{2,w}^{(t)} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & B_{w-1,w-1}^t & B_{w-1,w}^{(t)} \\ 0 & 0 & \cdots & 0 & B_{w,w}^t \end{pmatrix}. \tag{8}$$

Для установления её вида требуется определить вид блоков $B_{i,j}^{(t)}$ при $j > i$.

Рассмотрим блок B_{11}. Разобьём группу $\mathfrak{M_1}$ на подгруппы $(\mathfrak{M_{11}}, \mathfrak{M_{12}}, \mathfrak{M_{13}})$. К группе $\mathfrak{M_{12}}$ отнесём класс с номером из J, к группе $\mathfrak{M_{11}}$ — предшествующие ему классы, к группе $\mathfrak{M_{13}}$ — последующие классы. В соответствии с таким разбиением B_{11} принимает вид

$$B = \begin{pmatrix} C_{11} & C_{12} & 0 \\ 0 & C_{22} & C_{23} \\ 0 & 0 & C_{33} \end{pmatrix}, \tag{9}$$

А матрица B^t представляется в виде

$$B^t = \begin{pmatrix} C_{11}^t & C_{12}^{(t)} & C_{13}^{(t)} \\ 0 & C_{22}^t & C_{23}^{(t)} \\ 0 & 0 & C_{33}^t \end{pmatrix}. \tag{10}$$

Известно, что для неразложимой положительной матрицы A

$$A^t = uvr^t(1 + o(1)), \tag{11}$$

где r — перронов корень A, u и v — соответственно правый и левый собственные векторы, соответствующие r, причём $u > 0$, $v > 0$, $vu = 1$.

Таким образом, асимптотика матриц C_{11}^t, C_{22}^t, C_{33}^t известна.

Исследуем собственные векторы матрицы B_{11}, соответствующие числу r. Пусть $u = (u^{(1)}, u^{(2)}, u^{(3)})$ и $v = (v^{(1)}, v^{(2)}, v^{(3)})$ — соответственно правый и левый такие собственные векторы. Тогда

$$\begin{aligned} &C_{11}u^{(1)} + C_{12}u^{(2)} + C_{13}u^{(3)} = ru^{(1)} \\ &C_{22}u^{(2)} + C_{23}u^{(3)} = ru^{(2)} \\ &C_{33}u^{(3)} = ru^{(3)} \end{aligned} \quad . \tag{12}$$

Поскольку все собственные числа C_{33} строго меньше r, $u^{(3)} = 0$ и $u^{(2)}$ — правый собственный вектор C_{22}, относящийся к r, а $u^{(1)} = (rE - C_{11})^{-1}C_{12}u^{(2)}$.

Аналогично, рассматривая левый собственный вектор $v = (v^{(1)}, v^{(2)}, v^{(3)})$, имеем систему

$$\begin{aligned} &v^{(1)}C_{11} = rv^{(1)} \\ &v^{(1)}C_{12} + v^{(2)}C_{22} = rv^{(2)} \\ &v^{(1)}C_{13} + v^{(2)}C_{23} + v^{(3)}C_{33} = rv^{(3)} \end{aligned} \quad , \tag{13}$$

откуда $v^{(1)} = 0$, $v^{(2)}$ — левый собственный вектор C_{12}, и $v^{(3)} = v^{(2)}C_{23}(rE - C_{33})^{-1}$.

Выберем именно такие u и v, что $vu = 1$.

Рассмотрим асимптотику матрицы B_{11}^t. Нетрудно видеть, что

$$C_{12}^{(t)} = \sum_{i+j=t-1} C_{11}^i C_{12} C_{22}^j. \tag{14}$$

Разобьём эту сумму так, что $C_{12}^{(t)} = \Sigma_1 + \Sigma_2$, где

$$\begin{aligned} &\Sigma_1 : \begin{cases} i = \overline{t - \lfloor \log\log t \rfloor, t-1} \\ j = \overline{0, \lfloor \log\log t \rfloor - 1} \end{cases} \\ &\Sigma_2 : \begin{cases} i = \overline{0, t - \lfloor \log\log t \rfloor - 2} \\ j = \overline{\lfloor \log\log t \rfloor + 1, t-1} \end{cases} \end{aligned} \tag{15}$$

Рассмотрим вначале Σ_1. $C_{22}^t = u_2 v_2 r^t (1 + o(1)) \leq c_2$ при любых t. $C_{11}^t = u_1 v_1 (r')^t (1 + o(1)) \leq c_1 (r')^{t - \lfloor \log\log t \rfloor - 1}$. Отсюда,

$$\Sigma_1 \leq (r')^{t - \lfloor \log\log t \rfloor} \lfloor \log\log t \rfloor = O(\log\log t (\log t)^{c_3} (r')^t) = o(r^t). \tag{16}$$

Для Σ_2 $C_{22}^j = u_2 v_2 r^j (1 + o(1))$, поэтому $\Sigma_2 = \sum_{i+j=t-1} C_{11}^i C_{12} H r^j (1+o(1)) = r^{t-1} \sum_{i=0}^{t - \lfloor \log\log t \rfloor - 1} \left(\frac{C_{11}}{r}\right)^i C_{12} H (1 + o(1))$. Нетрудно видеть, что $\frac{1}{r} \sum_{i=0}^{\infty} \left(\frac{C_{11}}{r}\right)^i = (rE - C_{11})^{-1}$. Матрица $(rE - C_{11})^{-1}$ существует, так как все собственные числа C_{11} строго меньше r. Отсюда

$$\Sigma_2 = r^t (rE - C_{11})^{-1} C_{12} u^{(2)} v^{(2)} (1 + o(1)) = C_{12}^{(t)}. \tag{17}$$

Аналогично

$$C_{23}^{(t)} = \sum_{i+j=t-1} C_{22}^i C_{23} C_{33}^j, \tag{18}$$

откуда, проводя аналогичные вычисления, имеем

$$C_{23}^{(t)} = r^t u^{(2)} v^{(2)} C_{23} (rE - C_{33})^{-1} (1 + o(1)) \tag{19}$$

Подстановкой проверяется, что

$$C_{13}^{(t)} = \sum_{i+j=t-1} C_{12}^{(i)} C_{23} C_{33}^{j} \tag{20}$$

Подставляя в это выражение асимптотику для $C_{12}^{(t)}$, получаем:

$$C_{13}^{(t)} = r^t u^{(1)} v^{(2)} C_{23} (rE - C_{33})^{-1} (1 + o(1)) = r^t u^{(1)} v^{(3)} (1 + o(1)) \tag{21}$$

В результате, имеем:

$$B_{11} = \begin{pmatrix} 0 & u^{(1)} v^{(2)} & u^{(1)} v^{(3)} \\ 0 & u^{(2)} v^{(2)} & u^{(2)} v^{(3)} \\ 0 & 0 & 0 \end{pmatrix} r^t + o(r^t) \tag{22}$$

Получим теперь асимптотику всей матрицы A^t. Вначале пусть $w = 2$. Тогда

$$A^t = \begin{pmatrix} B_{11}^t & B_{12}^{(t)} \\ 0 & B_{22}^t \end{pmatrix} \tag{23}$$

Матрица B_{22}^t исследуется аналогично B_{11}^t, в результате имеем

$$B_{22}^t = \begin{pmatrix} u^{(22)} v^{(22)} & u^{(22)} v^{(32)} \\ 0 & 0 \end{pmatrix} r^t + o(r^t) \tag{24}$$

Для $B_{12}^{(t)}$ имеем

$$B_{12}^{(t)} = \sum_{i+j=t-1} B_{11}^{i} B_{12} B_{22}^{j} \tag{25}$$

Подставляя выражения для B_{11}^i и B_{22}^j, получаем:

$$B_{12}^{(t)} = \begin{pmatrix} 0 & u^{(1)} v^{(2)} & u^{(1)} v^{(3)} \\ 0 & u^{(2)} v^{(2)} & u^{(2)} v^{(3)} \\ 0 & 0 & 0 \end{pmatrix} \cdot B_{12} \cdot \begin{pmatrix} u^{(22)} v^{(22)} & u^{(22)} v^{(32)} \\ 0 & 0 \end{pmatrix} \cdot tr^t + o(tr^t) \tag{26}$$

Представляя B_{12} в блочном виде

$$B_{12} = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \\ D_{31} & D_{32} \end{pmatrix}, \tag{27}$$

и производя перемножение, получаем

$$B_{12}^{(t)} = \begin{pmatrix} u'^{(11)} v^{(22)} & u'^{(11)} v^{(32)} \\ u'^{(21)} v^{(22)} & u'^{(21)} v^{(32)} \\ 0 & 0 \end{pmatrix}. \tag{28}$$

Сформулируем теорему, определяющую вид блока $B_{lh}^{(t)}$ в общем случае.

Теорема 1

$$B_{lh}^{(t)} = u^{(l)} v^{(h)} t^{s_{lh}-1} r^t (1 + o(1)), \tag{29}$$

при $t \to \infty$, где $u^{(l)}$ и $v^{(h)}$ не зависят от t, и s_{lh} — число классов с номерами из J среди $K_l, K_{l+1}, \ldots, K_h$.

Доказательство. Доказательство проведём индукцией по w. При $w = 2$ теорема выполняется.

Пусть утверждение теоремы верно для $w - 1$ групп. Тогда

$$A = \begin{pmatrix} D_1 & E_1 \\ 0 & B_{w,w} \end{pmatrix} = \begin{pmatrix} B_{11} & E_2 \\ 0 & D_2 \end{pmatrix}, \tag{30}$$

где

$$E_1 = \begin{pmatrix} B_{1,w} \\ \vdots \\ B_{w-1,w} \end{pmatrix}, \quad E_2 = \begin{pmatrix} B_{12} & \cdots & B_{1,w} \end{pmatrix} \tag{31}$$

Тогда

$$A^t = \begin{pmatrix} D_1^t & E_1^{(t)} \\ 0 & B_{w,w}^t \end{pmatrix} = \begin{pmatrix} B_{11}^t & E_2^{(t)} \\ 0 & D_2^t \end{pmatrix} \tag{32}$$

Для матриц D_1, D_2 утверждение теоремы по индукции справедливо. Для доказательства теоремы достаточно рассмотреть

$$B_{1,w}^{(t)} = \sum_{l=1}^{w-1} \sum_{i+j=t-1} B_{1,l}^{(i)} B_{l,w} B_{w,w}^j \tag{33}$$

По предположению индукции слагаемое, содержащее $B_{1,w-1}^i = O(i^{s_{1,w-1}-1} r^i)$, преобладает над остальными. Поэтому

$$B_{1,w}^{(t)} = \sum_{i+j=t-1} B_{1,w-1}^{(i)} B_{w-1,w} B_{w,w}^j (1 + o(1)) \tag{34}$$

Подставляя по индукции выражение для $B_{1,w-1}$, получаем:

$$B_{1,w}^{(t)} = \begin{pmatrix} u'^{(11)} v^{(2w)} & u'^{(11)} v^{(3w)} \\ u'^{(21)} v^{(2w)} & u'^{(21)} v^{(3w)} \\ 0 & 0 \end{pmatrix} t^{s_{1,w}-1} r^t + o(t^{s_{1,w}-1} r^t) \tag{35}$$