О СВОЙСТВАХ ВЕРОЯТНОСТЕЙ ДЕРЕВЬЕВ ВЫВОДА ДЛЯ РАЗЛОЖИМОЙ СТОХАСТИЧЕСКОЙ КС-ГРАММАТИКИ

Л. П. Жильцова (Нижний Новгород)

В работе исследуются свойства вероятностей деревьев вывода высоты t при $t \to \infty$ для стохастической КС-грамматики с разложимой матрицей A первых моментов. Рассматривается критический случай, когда перронов корень матрицы A равен 1.

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, $R = \cup_{i=1}^{k} R_i$, где k - мощность алфавита V_N и R_i - множество правил с одинаковой левой частью A_i. Каждое правило r_{ij} из R_i имеет вид

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

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

Применение правила грамматики к слову состоит в замене вхождения нетерминала из левой части правила на слово, стоящее в его правой части.

Каждому слову α КС-языка соответствует последовательность правил грамматики (вывод), с помощью которой α выводится из аксиомы s. Выводу слова соответствует дерево вывода [1], вероятность которого определяется как произведение вероятностей правил, образующих вывод.

По стохастической КС-грамматике строится матрица A первых моментов. Для нее элемент a_l^i определяется как $\sum_j p_{ij} s_l^{(ij)}$, где величина $s_l^{(ij)}$ равна числу нетерминальных символов A_l в правой части правила r_{ij}. Перронов корень матрицы A обозначим через r.

Введем некоторые обозначения. Будем говорить, что нетерминал A_j непосредственно следует за нетерминалом A_i (и обозначать $A_i \to A_j$), если в грамматике существует правило вида $A_i \stackrel{p_{il}}{\to} \alpha_1 A_j \alpha_2$, где $\alpha_1, \alpha_2 \in (V_T \cup V_N)^*$. Рефлексивное транзитивное замыкание отношения $\to$ обозначим $\to_*$.

Классом нетерминалов назовем максимальное по включению подмножество $K \subseteq V_N$ такое, что $A_i \to_* A_j$ для любых $A_i, A_j \in K$. Для различных классов нетерминалов K_1 и K_2 будем говорить, что класс K_2 непосредственно следует за классом K_1 (и обозначать $K_1 \prec K_2$), если существуют $A_1 \in K_1$ и $A_2 \in K_2$, такие, что $A_1 \to A_2$. Рефлексивное транзитивное замыкание отношения $\prec$ обозначим через $\prec_*$.

Пусть $\{K_1, K_2, \ldots, K_m\}$ - множество классов нетерминалов грамматики, $m \geq 2$. Будем полагать, что классы нетерминалов перенумерованы таким образом, что $K_i \prec K_j$ тогда и только тогда, когда $i < j$.

Матрица первых моментов A имеет следующий вид:

$$A = \begin{pmatrix} A_{11} & A_{12} & \ldots & A_{1m-1} & A_{1m} \\ 0 & A_{22} & \ldots & A_{2m-1} & A_{2m} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & A_{m-1m-1} & A_{m-1m} \\ 0 & 0 & \ldots & 0 & A_{mm} \end{pmatrix}. \qquad (1)$$

Один класс нетерминалов представлен в матрице множеством подряд идущих строк и соответствующим множеством столбцов с теми же номерами. Для класса K_i квадратная подматрица, образованная соответствующими строками и столбцами, обозначается через A_{ii}. Подматрица A_{ij} является нулевой, если $K_i \not\prec K_j$. Блоки, расположенные ниже главной диагонали, нулевые в силу упорядоченности классов.

Для каждого класса K_i матрица A_{ii} неразложима. Без ограничения общности будем считать, что она строго положительна и непериодична. Обозначим через r_i перронов корень матрицы A_{ii}. Для неразложимой матрицы перронов корень является действительным и простым [2]. Очевидно, $r = \max_i\{r_i\}$.

Пусть $J = \{i_1, i_2, \ldots, i_l\}$ — множество всех номеров i_j классов, для которых $r_{i_j} = 1$.

Рассмотрим всевозможные последовательности классов $K_{i_1} \prec K_{i_2} \prec \ldots \prec K_{i_s}$, где $i_1 = 1, i_s = h$. Среди них выберем ту, которая содержит наибольшее число классов с номерами из J. Это число обозначим через s_{1h}. Дополнительно переупорядочим классы по неубыванию s_{1l}.

Рассмотрим также всевозможные последовательности классов $K_{i_1} \prec K_{i_2} \prec \ldots \prec K_{i_s}$, где $i_1 = i$. Среди них выберем ту, которая содержит наибольшее число классов с номерами из J. Это число обозначим через q_i.

Рассмотрим класс $K_j = \{A_1, A_2, \ldots, A_m\}$. Через $P_i^{(j)}(t)$ обозначим вероятность множества деревьев вывода высоты t, корень которых помечен нетерминалом A_i, и через $P^{(j)}(t) = \left(P_1^{(j)}(t), P_2^{(j)}(t), \ldots, P_m^{(j)}(t)\right)$ - вектор вероятностей для класса K_j.

ТЕОРЕМА 1. *Пусть* $r_j = 1$. *Тогда при* $t \to \infty$

$$P^{(j)}(t) \sim U^{(j)} \cdot \frac{c_j}{t^{1+\left(\frac{1}{2}\right)^{q_j-1}}},$$

где $U^{(j)}$ *- правый собственный вектор для матрицы* A_{ii} *из (1), соответствующий* r_j, *и* c_j *- некоторая константа.*

ТЕОРЕМА 2. *Пусть* $r_j < 1$. *Тогда при* $t \to \infty$

$$P^{(j)}(t) \sim U^{(j)} \cdot \frac{1}{t^{1+\left(\frac{1}{2}\right)^{q_j-1}}},$$

$$\text{где } U^{(j)} = (E - A_{jj})^{-1} \sum_{l} A_{jl} U^{(l)} \cdot c_l,$$

где суммирование ведется по всем номерам l *классов с* $r_l = 1$ *таких, что* $K_j \prec_* K_l$ *и* $q_l = q_j - 1$.

Здесь c_l и $U^{(l)}$ имеют тот же смысл, что и в теореме 1.

Из теорем 1 и 2 следует, что при $r_1 = 1$ наибольшую по порядку вероятность имеет множество деревьев вывода с корнем, помеченным нетерминалом из класса K_1.

При $r_1 < 1$ наибольшую по порядку вероятность имеют множества деревьев вывода с корнями, помеченными нетерминалами из классов K_l с $q_l = q_1$.

Список литературы

1.