ОБ АСИМПТОТИЧЕСКИХ ОЦЕНКАХ ДЛЯ БИГРАММНЫХ ЯЗЫКОВ

А. А. Петюшко (Москва)

Пусть A ($|A| < \infty$) — конечный алфавит.

Определение 1. *Биграммой* в алфавите A называется двухбуквенное слово $ab \in A^*, a, b \in A$ ($ab \neq ba$ при $a \neq b$).

Определение 2. Назовем *кратностью* β *в слове* α и обозначим через $\theta_\beta(\alpha)$, где $\beta \in A^*, \alpha \in A^*$, причем β — непустое слово, отображение $A^* \to N \cup \{0\}$, которое определяется как количество различных разложений слова α в виде $\alpha = \alpha'\beta\alpha''$ (α' и α'' могут быть пустыми). При длине слова α, меньшем чем длина слова β, значение $\theta_\beta(\alpha)$ положим равным 0.

С учетом введенных определений, по каждому слову $\alpha \in A^*$ можно построить квадратную матрицу биграмм $(\Theta(\alpha))_{i,j=1}^{|A|}$ размера $|A| \times |A|$ такую, что на месте (i, j) матрицы будет стоять значение $\theta_{a_i a_j}(\alpha)$ (при условии, что все буквы алфавита $A = \{a_1, a_2, ..., a_{|A|}\}$ пронумерованы и нумерация зафиксирована).

Обозначим через Ξ множество квадратных матриц размера $|A| \times |A|$, каждый элемент которых является неотрицательным целым числом. Т.о., $\forall \alpha \in A^*$ имеем $\Theta(\alpha) \in \Xi$. Также, здесь и далее через $\Theta(\alpha)$ будем обозначать матрицу биграмм, построенную по конкретному слову α, а через Θ - просто некоторую матрицу из Ξ, при этом будем считать, что на месте (i, j) матрицы Θ будет стоять значение $\theta_{a_i a_j}$.

Определение 3. Назовем *биграммным языком* $L(\Theta)$, порожденным матрицей $\Theta \in \Xi$, множество всех слов, имеющих одну и ту же матрицу биграмм Θ, т.е. $L(\Theta) = \{\beta \in A^* | \Theta(\beta) = \Theta\}$.

Построим по матрице $\Theta(\alpha)$ (или по произвольной матрице $\Theta \in \Xi$) ориентированный граф $G_{\Theta(\alpha)}$ на плоскости. Вершинами у этого графа будут все буквы из алфавита A, при этом ребра будут соответствовать биграммам с учетом их кратностей, т.е. кратность $\theta_{ab}(\alpha)$ будет порождать $\theta_{ab}(\alpha)$ ориентированных ребер $a \to b$. Аналогично, кратность $\theta_{cc}(\alpha)$ будет порождать $\theta_{cc}(\alpha)$ петель $c \to c$.

Определение 4. *Матрицей Кирхгофа* $ML(\Theta)$, построенной по матрице биграмм $\Theta \in \Xi$, называется квадратная матрица размером $|A| \times |A|$, т. ч. на месте (i, j) стоит элемент

$$l_{ij} = \begin{cases} -\theta_{a_i a_j}, & i \neq j; \\ \sum_{a_j \neq a_i} \theta_{a_i a_j}, & i = j. \end{cases}$$

Лемма. *Если матрица биграмм $\Theta \in \Xi$ такова, что соответствующий ориентированный граф G_Θ является эйлеровым [1], то все главные миноры $D^{(i,i)}$, полученные вычеркиванием из $ML(\Theta)$ i-й строки и i-го столбца, равны между собой при различных i (и равны D).*

На основе работы [2] можно доказать, что верна следующая

Теорема 1. *Пусть задана матрица биграмм Θ, которой соответствует эйлеров или почти эйлеров граф G_Θ, причем для $\forall i \exists j \neq i$, т.ч. $\theta_{a_i a_j} > 0$ или $\theta_{a_j a_i} > 0$. Тогда:*

1. *Если $\exists i'$, т.ч. $\sum_{a_i \in A} \theta_{a_i a_{i'}} > \sum_{a_i \in A} \theta_{a_{i'} a_i}$, то*

$$N_\Theta = \frac{\prod_{a_i \in A} (\sum_{a_j \in A} \theta_{a_i a_j} - 1 + \delta_{i' i})!}{\prod_{a_i, a_j \in A} \theta_{a_i a_j}!} D^{(i' i')};$$

где $\delta_{i' i}$ — символ Кронекера.

2. *Если $\forall i, j$ $\sum_{a_i \in A} \theta_{a_i a_j} = \sum_{a_i \in A} \theta_{a_j a_i}$, то*

$$N_\Theta = \left(\sum_{a_i, a_j \in A} \theta_{a_i a_j} \right) \frac{\prod_{a_i \in A} (\sum_{a_j \in A} \theta_{a_i a_j} - 1)!}{\prod_{a_i, a_j \in A} \theta_{a_i a_j}!} D.$$

Определение 5. Назовем *частотным языком* на биграммах с кратностями, заданным матрицей биграмм $\Theta \in \Xi$, следующий язык при $k \in N$:

$$F_\Theta = \bigcup_{k=1}^{\infty} L(k\Theta)$$

Определение 6. Назовем две ненулевые матрицы Θ_1 и Θ_2 из Ξ *кратными*, если существует действительный коэффициент $c \in R, c \neq 0$, такой что верно $\Theta_1 = c\Theta_2$. В противном случае ненулевые матрицы назовем некратными.

Теорема 2. *Пусть матрица биграмм Θ задает эйлеров граф G_Θ. Тогда:*

1. *Если существует такое разложение Θ в сумму двух ненулевых некратных матриц $\Theta = \Theta_1 + \Theta_2$ такое, что обе матрицы Θ_1 и Θ_2 задают эйлеровы графы G_{Θ_1} и G_{Θ_2}, то язык F_Θ нерегулярен.*

2. *В противном случае язык F_Θ регулярен.*

Определение 7. Матрица биграмм Θ называется *положительной*, если все элементы этой матрицы — натуральные целые числа.

Теорема 3. *Пусть задана положительная матрица биграмм* Θ *с эйлеровым графом* G_Θ. *Тогда при* $k \to \infty$ *мощность языка* $L(k\Theta)$

$$|L(k\Theta)| \cong c_2 * \frac{c_1^k}{k^{n(n-1)/2}},$$

где $c_1 = c_1(\Theta) > 1, c_2 = c_2(\Theta)$ *— некоторые константы, зависящие только от изначальной матрицы биграмм* Θ, *а* $n = |A|$ *— мощность алфавита.*

Обозначим через Ξ_k множество матриц размера $|A| \times |A|$, каждый элемент которых представляет собой неотрицательное целое число, не превосходящее $k > 0$. Также будем считать, что $|A| = n > 1$.

Через N_f^k обозначим количество матриц биграмм $\Theta \in \Xi_k$, задающих конечные (непустые) языки F_Θ; N_i^k — количество матриц биграмм $\Theta \in \Xi_k$, задающих счетные языки F_Θ; N_{reg}^k — количество матриц биграмм $\Theta \in \Xi_k$, задающих счетные регулярные языки F_Θ; N_{nreg}^k — количество матриц биграмм $\Theta \in \Xi_k$, задающих счетные нерегулярные языки F_Θ; N^k — общее количество матриц биграмм $\Theta \in \Xi_k$.

Теорема 4. *С учетом введенных выше обозначений верны следующие соотношения:*

1) $\exists k_0$, *т.ч.* $\forall k > k_0$ $\frac{1}{n(n-1)} < \frac{N_i^k}{N_f^k} < 1$;

2) $\lim_{k\to\infty} \frac{N_i^k}{N^k} = 0$;

3) $\lim_{k\to\infty} \frac{N_{reg}^k}{N_{nreg}^k} = 0$.

Следствие. *Если обозначить за* N_q^k *количество матриц биграмм* $\Theta \in \Xi_k$, *задающих непустые языки* F_Θ, *то* $\lim_{k\to\infty} \frac{N_q^k}{N^k} = 0$.

Список литературы

1. Оре О. Теория графов. — М.: Наука, 1980.

2. Hutchinson J. P., Wilf H. S. On Eulerian circuits and words with prescribed adjacency patterns // Journal of Combinatorial Theory. — 1975. — Ser. A, V. 18. — P. 80–87.