

Теорема 1. Пусть $p > 1$ — натуральное, а автомат $V' = (B, Q', \varphi', q'_0, Q'_F)$ представляет некоторый язык L . Тогда $L \in \mathcal{L}_p$ в том и только том случае, когда найдётся правильно p -раскрашенный автомат $V = (B, Q, \varphi, q_0, Q_F)$, также представляющий L . Причем число состояний нового автомата сравнимо с числом состояний исходного, а именно: $|Q| \leq \frac{p(p-1)}{2} \cdot |Q'| + p \cdot |Q'_F|$.

Теорема 2. Существует алгоритм, который для любого языка $L \in \text{Reg}$ определяет все такие p , что $L \in \mathcal{L}_p$.

Работа выполнена на кафедре МАТИС под руководством профессора д.ф.м.н. Бабина Д.Н.

Литература

- [1] Кудрявцев В. Б., Алешин С. В., Подколзин А. С. Введение в теорию автоматов. — М.: Наука, 1985.
- [2] Бабин Д. Н., Холоденко А. Б. Об автоматной аппроксимации естественных языков // Интеллектуальные системы. — 2007. — Т. 12 — С. 125–136.
- [3] Пархоменко Д. В Спектральная автоматная функция и связанные с нею автоматные языки // Интеллектуальные системы в производстве. — 2012. — Т. 9, № 1. — С. 165–175.

О биграммных языках с закольцовыванием

А. А. Петюшко

`petsan@newmail.ru`

МГУ им М. В. Ломоносова, Москва

Биграммные языки

Введение. Еще в начале 20 века выдающимся русским ученым А. А. Марковым был создан аппарат цепей, впоследствии названных цепями Маркова, и опробован [1] на вычислении переходных вероятностей между соседними буквами в поэме А. С. Пушкина „Евгений Онегин“. В дальнейшем этот аппарат получил широкое применение для распознавания и статистического моделирования естественных языков [2]. Тем не менее, в детерминированном случае, за редким исключением прикладных задач (например, для подсчета ДНК-последовательностей [3]), биграммы для исследования формальных языков практически не применялись. В данной статье автор изучает языки, состоящие из слов с фиксированными частотами пар соседних букв.

Определение биграммных языков. Пусть $A, |A| < \infty$ — конечный алфавит, A^* — множество всех конечных слов (включая пустое) в данном алфавите.

Определение 1. Биграммой в алфавите A называется двухбуквенное слово $ab \in A^*$, $a, b \in A$ (порядок вхождения букв в биграмму имеет значение, т.е. биграмма ab не равна биграмме ba при $a \neq b$).

Определение 2. Обозначим через $\theta_\beta(\eta)$, где $\beta \in A^*$, $\eta \in A^*$, причем β — непустое слово, отображение $A^* \rightarrow N \cup \{0\}$, сопоставляющее слову η число подслов β в слове η , т.е. количество различных разложений слова η в виде $\eta = \alpha'\beta\alpha''$ (α' и α'' могут быть пустыми). При длине слова η , меньшей длины слова β , значение $\theta_\beta(\eta)$ положим равным 0. Само же значение $\theta_\beta(\eta)$ при данных β и η назовем кратностью β в слове η .

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

Обозначим через Ξ множество квадратных матриц размера $|A| \times |A|$, каждый элемент которых является целым неотрицательным числом. Таким образом, для каждого $\alpha \in A^*$ 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 $\theta_{a_i a_j}$ (т.е. для произвольной матрицы из Ξ мы опустили зависимость от α как для самой матрицы биграмм, так и для отдельных ее элементов).

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. $L(\Theta) = \{\beta \in A^* | \Theta(\beta) = \Theta\}$.

Определение 4. Назовем частотным языком, заданным матрицей биграмм $\Theta \in \Xi$, язык $F_\Theta = \bigcup_{k=1}^{\infty} L(k\Theta)$, т.е. язык, состоящий из всех таких слов β , что набор кратностей этих слов $\Theta(\beta)$ кратен набору Θ , а именно, $F_\Theta = \{\beta \in A^* | \Theta(\beta) = k\Theta, k \in N\}$, где умножение k на Θ понимается как умножение скаляра на матрицу.

Подробно языки $L(\Theta)$ и F_Θ были изучены в [4]. В данной же работе изучаются свойства языков, в которых задающая их матрица биграмм учитывает так называемую „закольцованность“.

Биграммные языки с закольцовыванием

Определение 5. Элементарной матрицей кратностей биграмм $\Theta_{ij} \in \Xi$ будем называть матрицу из пространства матриц биграмм Ξ , имеющую единственный ненулевой элемент на месте (i, j) : $\theta_{a_i a_j} = 1, \theta_{a_k a_l} = 0, (k, l) \neq (i, j), 1 \leq i, j, k, l \leq n$.

Определение 6. Назовем $\Omega(\alpha)$ матрицей кратностей биграмм с закольцовыванием для непустого слова $\alpha \in A^*$ следующую матрицу:

- 1) при однобуквенном слове $\alpha = a_t, 1 \leq t \leq n, \Omega(\alpha) = \Theta_{tt}$,
- 2) при длине слова α не меньше 2, то есть $\alpha = a_i \alpha_1 a_j, 1 \leq i, j \leq n$, где $\alpha_1 \in A^*$ (в том числе α_1 может быть пустым), $\Omega(\alpha) = \Theta(\alpha) + \Theta_{ji}$.

Здесь и далее будем считать, что на месте (i, j) матрицы $\Omega(\alpha)$ будет стоять значение $\omega_{a_i a_j}(\alpha)$.

Содержательно, матрица биграмм с закольцовыванием — это та же обычная матрица биграмм за исключением единичной добавки в одной ячейке,

которая отвечает за биграмму, связывающую последнюю букву слова с первой (отсюда и название — „с закольцовыванием“, поскольку мы как бы считаем биграммы не на линейном слове, а на слове, начало и конец которого объединены в кольцо).

Определение 7. Назовем биграммным языком с закольцовыванием $K(\Omega)$ множество всех слов, имеющих одну и ту же матрицу Ω кратностей биграмм с закольцовыванием, т.е. $K(\Omega) = \{\beta \in A^* | \Omega(\beta) = \Omega\}$.

Рассмотрим основные свойства биграммных языков с закольцовыванием, которые во многом повторяют аналогичные утверждения для биграммных языков.

Построим по матрице $\Omega(\alpha)$ (или по произвольной матрице $\Omega \in \Xi$) ориентированный граф $G_{\Omega(\alpha)}$ (соответственно, G_{Ω}) 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 $\omega_{ab}(\alpha)$ будет порождать $\omega_{ab}(\alpha)$ ориентированных ребер $a \rightarrow b$. Аналогично, кратность $\omega_{cc}(\alpha)$ будет порождать $\omega_{cc}(\alpha)$ петель $c \rightarrow c$.

Также нам понадобятся понятия эйлеровых циклов и необходимых и достаточных условий существования оных в ориентированных графах из [5].

Теорема 1. Биграммный язык с закольцовыванием $K(\Omega)$ состоит не более чем из конечного числа слов одинаковой длины $l_{\Omega} = \sum_{a_i, a_j \in A} \omega_{a_i a_j}$. При этом язык $K(\Omega)$ непуст, если построенный по Ω ориентированный граф G_{Ω} — эйлеров.

Теорема 2. Пусть задана матрица $\Omega \in \Xi$ с эйлеровым графом G_{Ω} . Тогда существует взаимно-однозначное соответствие между словами языков $K(\Omega)$ и $L(\Theta)$, где $\Omega = \Theta$.

Следствие 1. Пусть для матрицы $\Omega \in \Xi$ ориентированный граф G_{Ω} — эйлеров. Тогда количество слов в языках $K(\Omega)$ и $L(\Theta)$, где $\Omega = \Theta$, одинаково: $|K(\Omega)| = |L(\Theta)|$.

Определение 8. Назовем частотным биграммным языком с закольцовыванием, заданным матрицей биграмм $\Omega \in \Xi$ с закольцовыванием, язык $E_{\Omega} = \bigcup_{k=1}^{\infty} K(k\Omega)$, т.е. язык, состоящий из всех таких слов β , т.ч. матрица биграмм $\Omega(\beta)$ с закольцовыванием этих слов кратна набору Ω , а именно $E_{\Omega} = \{\beta \in A^* | \Omega(\beta) = k\Omega, k \in N\}$, где умножение k на Ω понимается как умножение скаляра на матрицу.

Теорема 3. Если задана матрица биграмм $\Omega \in \Xi$ с закольцовыванием, то:
1) если ориентированный граф G_{Ω} является эйлеровым, то в частотном языке E_{Ω} с закольцовыванием счетное множество слов;
2) если ориентированный граф G_{Ω} не является эйлеровым, то в частотном языке E_{Ω} с закольцовыванием нет ни одного слова.

Определение 9. Назовем две ненулевые матрицы Ω_1 и Ω_2 из Ξ неколлинеарными, если не существует ненулевых действительных коэффициентов $c_1, c_2 \in R$, $(c_1, c_2) \neq (0, 0)$, для которых $c_1\Omega_1 + c_2\Omega_2 = 0$.

Теорема 4. Пусть задана матрица биграмм Ω с закольцовыванием такая, что ориентированный граф G_{Ω} — эйлеров. Тогда:

- 1) если существует такое разложение Ω в сумму двух ненулевых неколлинеарных матриц $\Omega = \Omega_1 + \Omega_2$ такое, что оба ориентированных графа G_{Ω_1} и G_{Ω_2} — эйлеровы, то частотный язык с закольцовыванием E_Ω нерегулярен;
- 2) в противном случае язык E_Ω регулярен. При этом для $\forall k \in N$ существуют ровно l слов $\beta_{k,i}$, $i = 1..l$, т.ч. $\Omega(\beta_{k,i}) = k\Omega$, а l — число ненулевых элементов в матрице Ω .

Литература

- [1] Марков А. А. Пример статистического исследования над текстом „Евгения Онегина“, иллюстрирующий связь испытаний в цепь // Известия Императорской Академии наук. — 1913. — Сер. 6, Т. 7, № 3. — С. 153–162.
- [2] Essen U., Steinbiss V. Cooccurrence smoothing for stochastic language modeling // IEEE International Conference on Acoustics, Speech, and Signal Processing. — 1992. — Vol. 1. — P. 161–164.
- [3] Hutchinson J. P., Wilf H. S. On eulerian circuits and words with prescribed adjacency patterns // Journal of Combinatorial Theory. — 1975. — Ser. A, Vol. 18. — P. 80–87.
- [4] Петлюшко А. А. О биграммных языках // Дискретная математика. — 2013. — Т. 25, № 3. — С. 64–77.
- [5] Оре О. Теория графов. — М.: Наука, 1980. — 336 с.

Решение проблемы эквивалентности и проблемы эквивалентных преобразований в одной двухпараметрической алгебраической модели программ

Р. И. Подловченко

podlovchenko.rimma@gmail.com

МГУ, Москва

Рассматриваемые нами алгебраические модели программ предназначены для изучения семантических свойств последовательных программ, использующих все стандартные композиции операторов, кроме аппарата процедур [1].

Сами программы, будучи записаны в формализованном виде, представляют собой конечный ориентированный граф со входом и выходом, все остальные вершины которого нагружены операторами и логическими условиями.

При моделировании программы объектом, называемым его схемой, сохраняется её структура, а операторы и логические условия заменяются соответственно операторными символами и логическими переменными, принимающими значения 0 и 1. Подходящая интерпретация тех и других возвращает схему в программу.

Алгебраическая модель программ состоит из множества схем программ, построенных над базисом операторных символов и логических переменных. Семантическая трактовка схемы связана с выполнением её на функциях разметки; такой функцией каждой операторной цепочке приписываются значения на ней всех логических переменных. Процедурой выполнения схемы на