

МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

DOI 10.17223/20710410/5/13

УДК 519.682

АНАЛИТИЧЕСКИЙ ПОДХОД В ТЕОРИИ КОНТЕКСТНО-СВОБОДНЫХ ЯЗЫКОВ В НОРМАЛЬНОЙ ФОРМЕ ГРЕЙБАХ¹

О. И. Егорушкин*, К. В. Сафонов**

** Красноярский государственный аграрный университет,**** Сибирский государственный аэрокосмический университет
имени академика М. Ф. Решетнева, г. Красноярск, Россия***E-mail:** safonovkv@rambler.ru

Контекстно-свободные языки рассматриваются как формальные степенные ряды, являющиеся решением системы полиномиальных уравнений с некоммутативными относительно умножения переменными. Предложено изучать эти системы в нормальной форме Грейбах, что позволит более эффективно использовать аналитические методы. Рассматриваются коммутативные образы контекстно-свободных языков и определяющих их систем уравнений в комплексной области.

Ключевые слова: *контекстно-свободные языки, нормальная форма Грейбах.*

Формальным языком L называют множество цепочек в алфавите $\{x_1, \dots, x_n\}$, выделенных с помощью конечного набора правил. Такие цепочки, принадлежащие свободной полугруппе $\{x_1, \dots, x_n\}^*$, называются грамматически правильными предложениями (над словарем), а конечное множество правил, с помощью которых выделяются эти цепочки, называют грамматикой. Таким образом, формальный язык определяется совокупностью соответствующих правил и способом выделения цепочек с помощью этих правил.

Практически важный класс формальных языков образуют контекстно-свободные языки (кс-языки), являющиеся мощным средством моделирования естественных языков и языков программирования [1–3]. Обозначим $W = \{x_1, \dots, x_n\} \cup \{z_1, \dots, z_m\}$ свободную полугруппу с операцией умножения над «расширенным» алфавитом $\{x_1, \dots, x_n, z_1, \dots, z_m\}$.

Словарем языка является конечное множество $X = \{x_1, \dots, x_n\}$ слов языка, называемое терминальным множеством, а $Z = \{z_1, \dots, z_m\}$ — множество вспомогательных символов z_j , необходимых для задания грамматических правил, называемое нетерминальным множеством; $W^* = (X \cup Z)^*$ — соответствующая свободная полугруппа относительно операции конкатенации. Дополним её операцией формального сложения «+» мономов из множества W^* (вместо суммы можно взять объединение « $\cup$ », следуя [1]), а также коммутативной операцией умножения мономов на (целые) числа. Таким образом, можно рассматривать не только многочлены, но и формальные степенные ряды с числовыми (как правило, целыми) коэффициентами от некоммутативных переменных.

¹Результаты работы докладывались на Международной конференции с элементами научной школы для молодёжи, г. Омск, 7–12 сентября 2009 г.

Кс-грамматика есть совокупность правил подстановки, которые каждому нетерминальному символу z_j ставят в соответствие некоторые мономы от терминальных и нетерминальных символов:

$$z_j \rightarrow f_{j1}(x, z), \dots, z_j \rightarrow f_{jq_j}(x, z),$$

причем z_1 — особый символ, играющий роль начала предложения в языке. Применение к нему произвольного числа подстановок порождает всевозможные правильные предложения языка.

Важной задачей теоретической информатики и ее приложений в теоретическом программировании является изучение структуры кс-языков и ее связи со структурой порождающих грамматик.

Правилам подстановки ставится в соответствие [3] система полиномиальных уравнений: каждому вспомогательному символу z_j , содержащемуся в левой части подстановки, сопоставляется уравнение $z_j = p_j(x, z)$, где $p_j(x, z) = f_{j1}(x, z) + \dots + f_{jq_j}(x, z)$. Таким образом, кс-грамматике соответствует система полиномиальных уравнений

$$z_j = p_j(x, z), j = 1, \dots, m, \quad (1)$$

называемая системой уравнений Хомского — Щютценберже, а кс-языком называется [3, 4] первая компонента z_1 ее решения $(z_1(x), \dots, z_m(x))$, получаемого методом последовательных приближений

$$z_j^{(k+1)} = p_j(x, z^{(k)}), z^{(0)} = (0, \dots, 0), j = 1, \dots, m,$$

где $z^{(k)} = (z_1^{(k)}, \dots, z_m^{(k)})$, 0 — нулевой моном, такой, что $0 \cdot u = u \cdot 0 = 0$ для любого монома u .

В результате итераций компоненты z_j выражаются формальными степенными рядами, причем кс-язык есть тот из них, который представляет выделенный символ z_1 :

$$z_1 = \sum_i < z_1, w_i > w_i. \quad (2)$$

Здесь $< z_1, w_i >$ — числовой коэффициент при мономе w_i от некоммутативных переменных $x_1, \dots, x_n$ (формальная сумма всех мономов и образует кс-язык). Условие того, что язык является контекстно-свободным, состоит также и в том, что многочлены $p_j(x, z)$ не содержат мономов z_j и e , где e — пустая цепочка. Таким образом, существование и единственность решения системы (1) обеспечивается её специфической структурой и методом последовательных приближений.

Система уравнений Хомского — Щютценберже впервые изучалась аналитическими методами в работах К. В. Сафонова (напр., [3]). Предложенный им подход состоял в изучении систем полиномиальных уравнений произвольной степени от многих переменных в комплексной области, что, естественно, сопряжено со значительными трудностями. В настоящей работе предлагается изучать вместо системы уравнений Хомского — Щютценберже эквивалентную ей систему в нормальной форме Грейбах [5, с. 127], в которой все многочлены имеют вторую степень по переменной z ; переход к нормальной форме Грейбах всегда возможен за счет увеличения, быть может, числа переменных. Для уравнений второй степени аналитические методы исследования представляются более перспективными: их эффективность в большей мере зависит от степени уравнений, чем от числа переменных.

Более точно, нормальная форма Грейбах системы (1) подразумевает, что все входящие в нее многочлены имеют вид $p_j(x, z) = f_j(z) + g_j(x, z) + h_j(x)$, где $f_j(z)$ — квадратичная форма от переменных $z_1, \dots, z_m$, многочлен $g_j(x, z)$ линеен по $z_1, \dots, z_m$, а многочлен $h_j(x)$ от них не зависит [5, с. 127]. А значит, можно считать, что система (1) имеет вид

$$z_j = f_j(z) + g_j(x, z) + h_j(x) = 0, j = 1, \dots, m. \quad (1^*)$$

Нам понадобится понятие коммутативного образа формального ряда (многочлена) [2, 3]. А именно, поставим в соответствие формальному степенному ряду (многочлену) ряд (многочлен) с комплексными переменными, задав отображение терминальных x_i и нетерминальных z_i символов из множества $X \cup Z$ в множество комплексных переменных, оставив за ними прежние обозначения, тогда далее $(x, z) \in C_{x,z}^{m+n}$. Таким образом, получаем фиксированный гомоморфизм, который ставит в соответствие формальному ряду r его *коммутативный образ* — сходящийся в окрестности нуля степенной ряд от комплексных переменных

$$ci(r) = \sum_k a_k x^k, \quad (3)$$

где

$$a_k x^k = a_{k_1, \dots, k_n} x_1^{k_1} \dots x_n^{k_n},$$

$$a_k = \sum_{\#x_1(w_i)=k_1, \dots, \#x_n(w_i)=k_n} < r, w_i >, \quad (4)$$

символ $\#c(d)$ означает число вхождений символа c в моном d .

Итак, рассмотрим систему уравнений (1^*) . Если в ней $f_j(z) \equiv 0$ при всех j , то эта система линейная и порождаемый ею язык принадлежит к простейшему классу линейных языков [1, 2]. Однако важно, что существует связь между кс- и линейными языками, точнее, между их коммутативными образами.

Рассмотрим коммутативный образ кс-языка z_1

$$ci(z_1) = \sum a_k x^k = \sum a_{k_1, \dots, k_n} x_1^{k_1} \dots x_n^{k_n}, \quad (5)$$

который является алгебраической функцией, а также коммутативный образ

$$ci(L) = \sum_{k_0 \geq 0, k \geq 0} L_{k_0, k} x_0^{k_0} x^k \quad (6)$$

некоторого линейного языка над терминальными символами $x_0, x_1, \dots, x_n$, который является рациональной функцией.

Будем называть ряд (5) *диагональю* ряда (6), если при всех $k_1, \dots, k_n$ выполнено условие

$$a_{k_1, \dots, k_n} = L_{k_1, k_1, k_2, \dots, k_n},$$

и в следующей теореме установим, для каких именно кс-языков коммутативные образы являются диагоналями линейных языков.

Теорема 1. Если однородные многочлены второй степени $f_2(z), \dots, f_m(z)$, входящие в систему (1^*) , не зависят от переменной z_1 и система уравнений

$$f_j(z) = 0, j = 2, \dots, m, \quad (7)$$

имеет единственный нуль $z = 0$, то коммутативный образ кс-языка, порожденного системой (1^*) , является диагональю некоторого линейного языка.

Учитывая относительную простоту структуры линейных языков и конструкции диагонали, можно сделать вывод о том, что структура широкого класса кс-языков, описанных в теореме 1, является достаточно простой.

Доказательство. Рассмотрим корни системы (1^*) следующим образом: будем считать переменные x и z_1 комплексными параметрами, а остальные корни $z[1] = (z_2, \dots, z_m) = z[1](x, z_1)$ зависящими от этих параметров. Условие теоремы 1 означает, что ни при каких значениях параметров корни системы (1^*) не стремятся к бесконечности (как это бывает, например, когда старший коэффициент многочлена при некотором значении параметра обращается в нуль). Следовательно, при всех значениях параметров x и z_1 число корней $z^{(k)}[1](x, z_1)$ подсистемы $p_j(x, z) = 0, j = 2, \dots, m$, с учетом кратностей μ_k одно и то же, а потому возможно корректно определить результат этой подсистемы относительно уравнения $p_1(x, z) = 0$:

$$P(x, z_1) = \prod_k p_1^{\mu_k}(x, z_1, z^{(k)}[1](x, z_1)),$$

где в произведение входят все корни уравнения. В силу симметричной зависимости этой функции от корней и алгебраичности самих корней результат также является симметричной алгебраической функцией, т. е. многочленом.

Согласно теореме о результате системы [6, с. 238], кратность корня $(0, 0)$ системы (1^*) , равная 1, совпадает с кратностью корня $z_1 = 0$ многочлена $P(x, z_1)$ при $x = 0$. А значит, для многочлена $P(x, z_1)$ выполняется условие

$$P(0, 0) = 0, \frac{\partial P(0, 0)}{\partial z_1} \neq 0.$$

Отсюда следует, что алгебраическая функция z_1 , определяемая многочленом P , является диагональю некоторой рациональной функции от переменных $x_0, x_1, \dots, x_n$ (см. [7]). Поскольку каждая голоморфная в нуле рациональная функция является коммутативным образом соответствующего линейного языка, теорема 1 доказана. ■

Порождающая кс-язык система уравнений (1^*) в нормальной форме Грейбах позволяет установить и другие свойства кс-языков, например, является ли данный формальный степенной ряд контекстно-свободным языком или нет. Сгруппируем коммутативный образ формального языка в ряд по однородным многочленам:

$$ci(z_1) = \sum_j (a)_j(x), \quad (8)$$

где $(a)_j(x)$ — однородный многочлен степени j . Достаточно установить, что ряд (8) удовлетворяет уравнению степени 2, поскольку такую степень имеет система (1^*) . Возводя ряд (8) в квадрат, получим ряд $(ci(z_1))^2 = \sum_j (a)_j^{(2)}(x)$. Тот факт, что оба ряда, будучи умноженными на многочлены, дают в сумме тождественный нуль, означает, что достаточно длинные отрезки этих рядов линейно зависимы. Таким образом, получается следующая теорема.

Теорема 2. Для того чтобы ряд по однородным многочленам (8) удовлетворял полиномиальному уравнению степени 2, необходимо и достаточно, чтобы при всех

$j \geq j_0, l \geq l_0$ выполнялось равенство

$$\begin{vmatrix} (a)_j(x) & \dots & (a)_{j+l}(x) & (a)_j^{(2)}(x) & \dots & (a)_{j+l}^{(2)}(x) \\ (a)_{j+1}(x) & \dots & (a)_{j+l+1}(x) & (a)_{j+1}^{(2)}(x) & \dots & (a)_{j+l+1}^{(2)}(x) \\ \dots & \dots & \dots & \dots & \dots & \dots \\ (a)_{j+q}(x) & \dots & (a)_{j+l+q}(x) & (a)_{j+q}^{(2)}(x) & \dots & (a)_{j+l+q}^{(2)}(x) \end{vmatrix} \equiv 0,$$

где $q = 2l + 1$.

Таким образом, можно сделать вывод о том, что нормальная форма Грейбах систем уравнений, порождающих кс-языки, имеет ряд преимуществ, связанных с тем, что эти уравнения имеют степень не выше второй. Так, нормальная форма Грейбах позволяет установить важную связь кс-языков с линейными языками (теорема 1), а также получить фундаментальное условие, характеризующее кс-языки (теорема 2).

ЛИТЕРАТУРА

1. Глушков В. М., Цейтлин Г. Е., Ющенко Е. Л. Алгебра, языки, программирование. Киев: Наук. думка, 1974. 328 с.
2. Семенов А. Л. Алгоритмические проблемы для степенных рядов и контекстно-свободных грамматик // Докл. АН СССР. 1973. Т. 212. С. 50–52.
3. Сафонов К. В. О возможности вычислительного распознавания контекстно-свободных языков // Вычислительные технологии. 2005. Т. 10. № 4. С. 91–98.
4. Сафонов К. В., Егорушкин О. И. О синтаксическом анализе и проблеме В. М. Глушкова распознавания контекстно-свободных языков Хомского // Вестник Томского государственного университета. Приложение. 2006. № 17. С. 63–66.
5. Salomaa A., Soittola M. Automata-Theoretic Aspects of Formal Power Series. N.Y.: Springer Verlag, 1978. 176 p.
6. Айзенберг Л. А., Южаков А. П. Интегральные представления и вычеты в многомерном комплексном анализе. Новосибирск: Наука, 1979. 366 с.
7. Safonov K. V. On Power Series of Algebraic and Rational Functions in C^n // J. Mathematical Analysis and Applications. 2000. V. 243. P. 261–277.