УДК 519.71

Применение сетей Петри для анализа КС-грамматик

И.Я.Спекторский

Предлагается схема использования сетей Петри для исследования некоторых свойств КС-грамматик. Метод позволяет, в частности, исследовать заданную КС-грамматику на пустоту и конечность порождаемого языка, используя дерево покрываемости соответствующей сети Петри. Кроме того, предложенный метод позволяет сформулировать необходимые условия порождения заданного слова КС-грамматикой в терминах матричного анализа соответствующей сети.

Пропонується схема застосування мереж Петрі для дослідження деяких властивостей КВ-граматик. Метод дозволяє, зокрема, досліджувати задану КВ-граматику на порожність та скінченність породжуваної мови, використовуючи дерево покриття відповідної мережі Петрі. Крім того, запропонований метод дозволяє сформулювати необхідні умови породження заданого слова КВ-граматикою у термінах матричного аналізу відповідної мережі.

The paper proposes the method of analyzing some properties of context-free grammars with a help of Petri net. This method, in particular, enables to analyze the emptiness and finiteness of language, generated by given CF-grammar, using coverability tree of the respective Petri net. Additionally, the proposed method can be applied to form some necessary conditions for given string to be generated by CF-grammar within matrix state equation of the Petri net.

Введение

В работах [1-3] описан следующий метод представления формальных языков с помощью порождающих сетей Петри: каждому переходу сети сопоставляется либо один из символов терминального алфавита, либо «пустой символ» λ, и каждая последовательность запусков переходов сети, заканчивающаяся в одной из выделенных «терминальных» маркировок, определяет слово порождаемого языка. В зависимости от множества терминальных маркировок и наличия λ-переходов определяют более десяти различных классов языков, порождаемых сетями Петри (см., напр., [1]), которые не вписываются в классическую иерархию Хомского – строго включают класс регулярных языков, строго вложены в класс контекстно-зависимых языков, и несравнимы с классом контекстно-свободных языков (детально теория формальных языков и грамматик изложена, напр., в [4-6]).

Метод анализа, предложенный в этой статье, не предполагает полного описания языка сетью Петри, и поэтому, в частности, не дает критерия порождаемости данного слова заданной формальной грамматикой, но позволяет с помощью сети Петри контролировать количество вхождений каждой буквы терминального алфавита в порождаемое слово и, как следствие, исследовать порождаемый язык на пустоту и конечность.

Определение сети Петри для заданной КС-грамматики

Пусть G=<N, Σ, P, S> - контекстно-свободная порождающая грамматика (КС-грамматика), где N - нетерминальный алфавит, Σ - терминальный алфавит, P - множество продукций, $S \in N$ - источник. Для грамматики G введем в рассмотрение сеть Петри N_G с множеством позиций $N \cup \Sigma$, множеством переходов P и весовой функцией W, определяемой для продукции $A \to \beta$ и символа $\xi \in \Sigma$ соотношениями:

- $W(A \to \beta, \xi) = |\beta|_\xi$ (количество вхождений ξ в β).
- $W(\xi, A \to \beta) = \begin{cases} 1, \text{если } \xi = A, \\ 0, \text{если } \xi \neq A. \end{cases}$

Пример 1.

Рассмотрим формальную грамматику G=<{S,A}, {a,b,c}, P, S>, где множество продукций P={$S \to aSab \mid A,\ A \to cA \mid \varepsilon$}. Сеть Петри, соответствующая грамматике G, изображена на рис. 1.

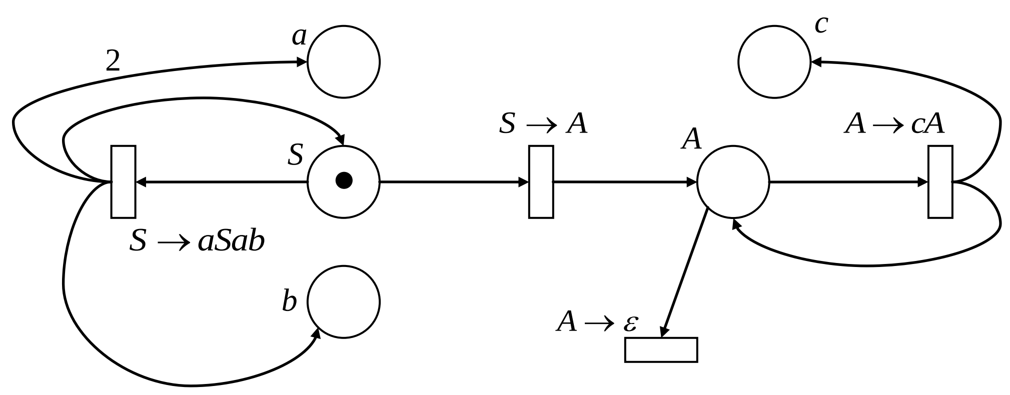

Рис. 1. Сеть Петри для грамматики с продукциями $S \to aSab \mid A,\ A \to cA \mid \varepsilon$.

Легко видеть, что данная грамматика порождает язык $\{a^n c^m (ab)^n : n \geq 0, m \geq 0\}$, причем каждое слово $a^n c^m (ab)^n$ порождается последовательностью применений продукций (запусков переходов) $(S \to aSab)^n (S \to A)(A \to cA)^m (A \to \varepsilon)$, что приводит к маркировке (0,0,n,n,m) (предполагая порядок позиций в соответствии с перечислением в определении грамматики, т.е. S,A,a,b,c).

Анализ с помощью дерева покрываемости

Ряд свойств сети Петри N_G при фиксированной начальной маркировке ($\mathbf{N} = \{1,2,\ldots,n,\ldots\}$ - множество натуральных чисел) эффективно анализируются с помощью дерева покрываемости (см., напр. [1,2]).

Символ $A \in N$ называют *порождающим*, если $A \overset{*}{\Rightarrow} w$ для некоторого $w \in \Sigma$. Очевидно, что все продукции, содержащие непорождающие символы, можно исключить из P, не сужая язык, порождаемый грамматикой. Продукцию $A \to \beta$ называют *рекурсивной*, если β содержит A. Очевидно, что исключение рекурсивных продукций не влияет на порождаемость или непорождаемость нетерминальных символов.

Через μ_A ($A \in N$) обозначим такую маркировку, что

$$\mu_A = \begin{cases} 1, \text{если } \xi = A, \\ 0, \text{если } \xi \in (N \cup \Sigma) \setminus \{A\}. \end{cases}$$

Через T_A обозначим дерево покрываемости для начальной маркировки μ_A. Через T_A^{nr} обозначим дерево покрываемости для начальной маркировки μ_A, построенное без учета переходов, соответствующих рекурсивным продукциям.

Справедливость следующих двух теорем устанавливается непосредственно из построения сети Петри по заданной КС-грамматике..

Теорема 1. Символ $A \in N$ является порождающим тогда и только тогда, когда дерево T_A^{nr} содержит не менее одной маркировки μ, такой, что $\mu(\xi)=0$ для всех $\xi \in N$.

Следствие. Язык, порождаемый грамматикой G, является непустым тогда и только тогда, когда дерево T_S^{nr} содержит не менее одной маркировки μ, такой, что $\mu(\xi)=0$ для всех $\xi \in N$ (т.е., когда порождающим является источник S).

Теорема 2. Грамматика G порождает бесконечный язык тогда и только тогда, когда дерево T_S содержит не менее одной маркировки μ, такой, что $\mu(\xi)=0$ для всех $\xi \in N$, и $\mu(a)=\omega$ для некоторого $a \in \Sigma$.

Пример 2. Рассмотрим формальную грамматику $G=<\{S,A\}, \{a,b,c\}, P, S>$, где множество продукций $P=\{ S \to aSab \mid cA,\ A \to c \mid \varepsilon \}$. Сеть Петри, соответствующая грамматике G, изображена на рис. 2.

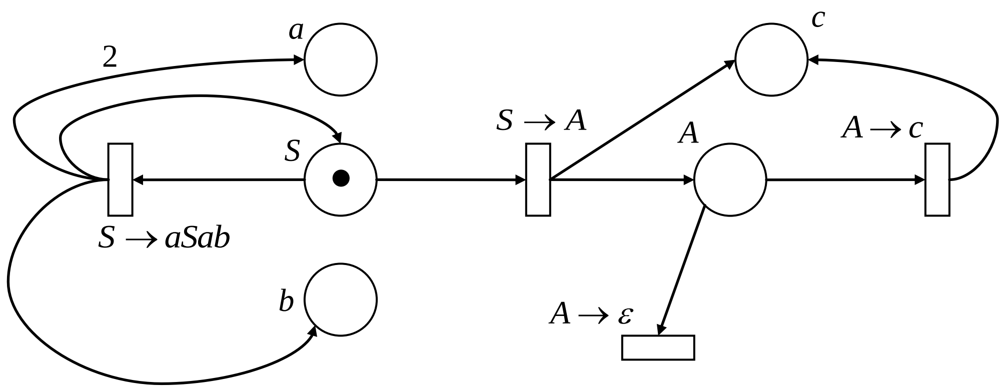

Рис. 2. Сеть Петри для грамматики с продукциями $S \to aSab \mid cA,\ A \to c \mid \varepsilon$.

Для анализа порождаемого языка на пустоту и конечность построим дерево T_S^{nr} (см. рис. 3) и дерево T_S (см. рис. 4). При построении дерева T_S^{nr}, в соответствии с определением, была исключена рекурсивная продукция $S \to aSab$. Порядок позиций при записи маркировок предполагался в соответствии с перечислением в определении грамматики, т.е. *S*,*A*,*a*,*b*,*c*.

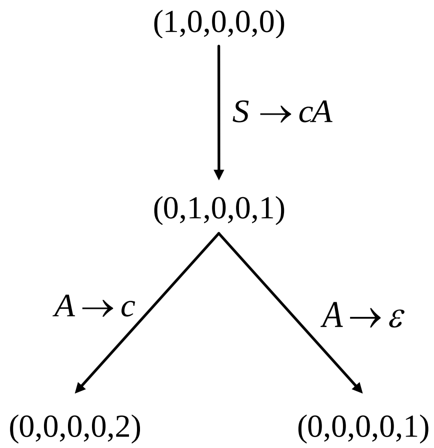

Рис. 3. Дерево T_S^{nr} для грамматики с продукциями $S \to aSab \mid cA,\ A \to c \mid \varepsilon$.

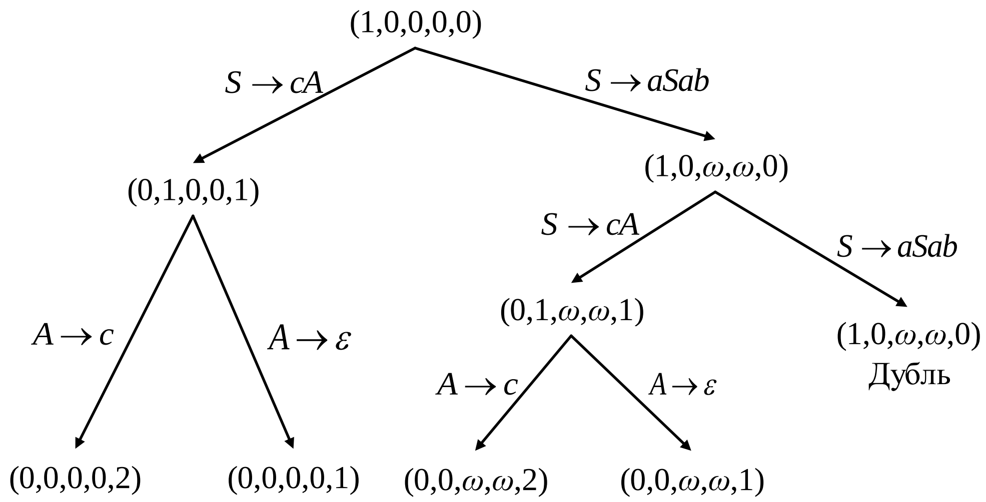

Рис. 4. Дерево T_S для грамматики с продукциями $S \to aSab \mid cA,\ A \to c \mid \varepsilon$.

Поскольку T_S^{nr} содержит маркировки μ, удовлетворяющие условию μ(ξ)=0 для всех $\xi \in \{S,A\}$ (например, маркировку (0,0,0,0,2)), заданная грамматика порождает непустой язык. Далее, дерево T_S содержит маркировки μ, удовлетворяющие условию $\mu(\xi)$=0 для всех $\xi \in \{S,A\}$ и содержащие символ ω (маркировки (0,0,ω,ω,1) и (0,0,ω,ω,2)), откуда следует бесконечность языка, порождаемого данной грамматикой. Более того, поскольку символ ω находится на позициях, соответствующих символам a и b, можем сделать вывод, что порождаемый язык содержит слова с бесконечным количеством букв a и b (но не c).

Анализ с помощью уравнения состояния

Для анализа порождаемости грамматикой G слова $w \in \Sigma^*$ можно использовать уравнение состояний сети N_G (см., напр., [1,2]). Обозначим через μ_w ($w \in \Sigma^*$) маркировку, такую, что $\mu_w(\xi) = |w|_\xi$. Очевидно, что для порождаемости слова w необходимо (но недостаточно), чтобы уравнение состояния имело хотя бы одно решение для правой части $\Delta\mu = \mu_w - \mu_S$.

Пример 3.

Рассмотрим грамматику G=<{S}, {a,b}, {$S \to aSb \mid bSa \mid \varepsilon$}, S>. Соответствующая сеть Петри N_G изображена на рис. 5.

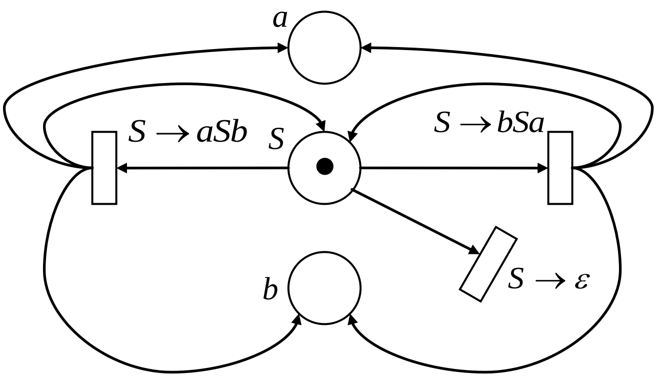

Рис. 5. Сеть Петри для грамматики с продукциями $S \to aSb \mid bSa \mid \varepsilon$.

Выпишем матрицу инцидентности для N_G, предполагая упорядоченность позиций (нетерминальных и терминальных символов) и переходов (продукций) в соответствии с перечислением в определении грамматики:

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 0 \end{pmatrix}.$$

Уравнение состояния $A^T u = \Delta\mu$ имеет решение относительно вектора запусков $u = (u_1, u_2, u_3)$ лишь для правых частей вида $\Delta\mu = (x, y, y)$. Таким образом, из начальной маркировки $\mu_S = (1,0,0)$ могут достигаться только маркировки с одинаковыми второй и третьей координатами. Это означает, что грамматика *G* может генерировать лишь слова с одинаковым числом вхождений *a* и *b*. Однако разрешимость уравнения состояния - лишь необходимое условие для порождаемости слова грамматикой; в данном примере *G* порождает лишь слова вида ww^r, т.е. палиндромы четной длины.

Литература

1. Питерсон Джеймс. Теория сетей Петри и моделирование систем / М.: Мир, 1984. – 264 с.

2. Котов Вадим Евгеньевич. Сети Петри / М.: Наука, 1984. – 160 с.

3. Алгоритмічні алгебри : навч. посібник / К.: ІЗМН, 1997. – 480 с. – ISBN 5-7763-9094-X.

4. Рейуорд-Смит В. Дж. Теория формальных языков. Вводный курс / М.: Радио и связь, 1988. – 128 с. – ISBN 5-256-00159-0.

5. Гросс М., Лантен А. Теория формальных грамматик / М.: Мир, 1971. – 296 с.

6. Ахо А., Ульман Дж. Теория синтаксического анализа, перевода и компиляции. Т. 1 / М.: Мир, 1978. – 614 с.