Доклады Академии наук СССР
1970. Том 192, № 5

МАТЕМАТИКА

А. Я. ДИКОВСКИЙ

ЯЗЫКИ ОГРАНИЧЕННОЙ АКТИВНОЙ ЕМКОСТИ

(Представлено академиком П. С. Новиковым 15 XII 1969)

В этой заметке исследуются структурные свойства языков, порождаемые грамматиками с некоторыми ограничениями на «рабочую емкость» вывода.

Мы будем рассматривать порождающие грамматики с правилами вида $\omega_1\varphi\omega_2 \to \omega_1\psi\omega_2$, где ω_1, ω_2 — цепочки в основном алфавите, а φ — непустая цепочка, самый левый и самый правый символы которой являются вспомогательными (в частном случае сюда попадают часто рассматриваемые в литературе порождающие грамматики с правилами вида $\varphi \to \psi$, где φ — непустая цепочка вспомогательных символов).

Пусть $\Gamma = (V, V_1, I, R)$ — некоторая грамматика рассматриваемого класса. Активной емкостью цепочки φ в алфавите $V \cup V_1$ (обозначение $|\varphi_1|$) мы будем называть количество вхождений в φ вспомогательных символов, а активной емкостью вывода $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ в Γ — число $s = \max\limits_{1 \leqslant i \leqslant n} \{|\varphi_i|_1\}$. Активной емкостью цепочки $x \in L(\Gamma)$ (обозначение $s_{\inf}(x)$) мы назовем наименьшую активную емкость полного вывода x в Γ (т. е. вывода x из начального символа I). Если найдется число k такое, что для каждой цепочки $x \in L(\Gamma)$ $s_{\inf}(x)$ не превышает k, мы будем говорить, что Γ является грамматикой ограниченной активной емкости; при этом наименьшее из всех указанных чисел k мы будем называть активной емкостью Γ и обозначать $s_{\inf}(\Gamma)$. Нас будет интересовать и случай, когда существует число l, ограничивающее сверху активную емкость любого полного вывода в Γ; тогда мы будем говорить, что Γ является грамматикой равномерно ограниченной активной емкости, а наименьшее из всех указанных чисел l будем называть предельной активной емкостью Γ и обозначать $s_{\sup}(\Gamma)$.

Теорема 1. *По всякой грамматике Γ ограниченной (равномерно ограниченной) активной емкости можно построить эквивалентную ей кс-грамматику Γ' ограниченной (равномерно ограниченной) активной емкости, причем $s_{\inf}(\Gamma') \leqslant s_{\inf}(\Gamma)$ (соответственно $s_{\sup}(\Gamma') \leqslant s_{\sup}(\Gamma)$).*

Теорема 1 позволяет нам без потери общности ограничиться классом кс-грамматик ограниченной активной емкости.

Удобным аппаратом для исследования кс-грамматик ограниченной активной емкости оказывается следующая мера сложности дерева вывода — мы будем называть ее густотой и обозначать μ.

Пусть $\Gamma = (V, V_1, I, R)$ — произвольная кс-грамматика и γ — дерево полного вывода в Γ цепочки x (обозначение $\gamma = (I \underset{\Gamma}{\leqslant} x)$).

1. Если α — тупиковая вершина γ, то $\mu(\alpha) = 0$.
2. Пусть из α дуги ведут в вершины $\alpha_1, \alpha_2, \ldots, \alpha_r$; тогда $\mu(\alpha) = 1 + \max\limits_{1 \leqslant l \leqslant r} \{\mu(\alpha_l)\}$, если найдутся i и j, $1 \leqslant i \neq j \leqslant r$, такие, что $\mu(\alpha_i) = \mu(\alpha_j) = \max\limits_{1 \leqslant l \leqslant r} \{\mu\alpha_l)\}$, и $\mu(\alpha) = \max\limits_{1 \leqslant l \leqslant r} \{\mu(\alpha_l)\}$ в противном случае. Густота $\mu(\gamma)$ дерева вывода γ есть густота $\mu(\beta)$ его корня β.

Густота $\mu(\gamma)$ дерева полного вывода γ оказывается очень просто связанной с его активной емкостью $s_{\inf}(\gamma)$ (т. е. с наименьшей активной емкостью вывода, производимого деревом γ).

Теорема 2. *Для всякой кс-грамматики* Γ, *максимальная длина правых частей правил которой равна* d, *и для всякого дерева полного вывода* γ *в* Γ $\mu(\gamma) \leqslant s_{\inf}(\gamma) \leqslant d[\mu(\gamma)+1]$.

Теорема 2 дает простой критерий ограниченности активной емкости кс-грамматики Γ, состоящий в том, что найдется число c такое, что для всякой цепочки $x \in L(\Gamma)$ существует дерево вывода $\gamma = (I \underset{\Gamma}{\leqslant} x)$ с густотой $\mu(\gamma) \leqslant c$.

Э. Д. Стоцкий (1) поставил вопрос о соотношении между классом всех кс-языков и класом языков ограниченной активной емкости, т. е. языков, порождаемых кс-грамматиками ограниченной активной емкости. Этот вопрос решает

Теорема 3. *Никакая кс-грамматика ограниченной активной емкости не порождает язык* L_0 *всех бинарных скобочных последовательностей* (*т. е. язык*: а) *содержащий последовательность* ( ), б) *содержащий последовательность* $(z'z'')$ *для любых последовательностей* z' *и* z'', *ему принадлежащих,* в) *не содержащий никаких других последовательностей*).

Интересной особенностью языков ограниченной активной емкости является то, что они «строятся» из линейных языков с помощью операции подстановки (отображение h языка в алфавите $V=\{a_1,\ldots,a_n\}$ на некоторый другой язык называется подстановкой, если для каждого i, $1 \leqslant i \leqslant n$, $h(a_i)$ есть некоторый язык и $h(a_{i_1}\ldots a_{i_s}) = h(a_{i_1})\ldots h(a_{i_s})$. Точнее, имеет место

Теорема 4. *Класс языков ограниченной активной емкости совпадает с наименьшим классом языков, содержащим все линейные языки и замкнутым относительно операции подстановки.*

При доказательстве этой теоремы существенным образом используется теорема 2.

Дальнейшие результаты относятся к языкам, порождаемым кс-грамматиками равномерно ограниченной активной емкости (языкам равномерно ограниченной активной емкости).

Вопрос о соотношении между классами языков ограниченной и равномерно ограниченной активной емкости решается так.

Теорема 5. *Для любого числа* $k = 1, 2, \ldots$ *язык* $\{a^n b^n \mid n = 1, 2, \ldots\}^k$ *порождается подходящей кс-грамматикой* Γ *с предельной активной емкостью* k, *но не порождается никакой кс-грамматикой* Γ' *с предельной активной емкостью, меньшей* k.

Из этой теоремы вытекает очевидное

Следствие. *Язык* $L_1 = \bigcup\limits_{j=1}^{\infty} \{a^n b^n \mid n = 1, 2, \ldots\}^j$ *не порождается никакой кс-грамматикой равномерно ограниченной активной емкости.*

Остается заметить, что язык L_1 порождается кс-грамматикой с активной емкостью два.

У языков равномерно ограниченной активной емкости имеется структурная характеристика, подобная той, которая была получена выше для класса всех языков ограниченной активной емкости. Как оказалось, языки этого класса строятся из линейных языков с помощью операций объединения, умножения и подстановки в центрированный линейный язык (линейный язык L в алфавите $V = V' \cup \{C\}$, $C \notin V'$, является центрированным, если $L \subseteq V'^* c V'^*$, c называется центром языка L; h есть подстановка в центрированный линейный язык L, если h переводит центр L в некоторый язык, а на остальных символах алфавита является тождественным). В точной форме это утверждение выглядит так:

Теорема 6. *Класс языков равномерно ограниченной активной емкости совпадает с наименьшим классом языков, содержащим все линейные языки и замкнутым относительно операций объединения, умножения и подстановки в центрированный линейный язык.*

В заключение приведем еще один результат, по форме несколько выпадающий из нашего изложения, но тоже дающий некоторую характеристику класса языков равномерно ограниченной активной емкости.

Пусть $\Gamma = (V, V_1, I, R)$ — порождающая грамматика указанного выше вида. Сопоставим каждому правилу из R некоторый символ r так, чтобы разным правилам соответствовали различные пометки; множество всех пометок обозначим через $\bar{R}$. Тогда каждому полному выводу D в Γ естественным образом приписывается цепочка над алфавитом $\bar{R}$ — характеристика D, а грамматике Γ соответствует множество всех характеристик полных выводов в Γ — надстройка Γ.

Теорема 7. *Язык L тогда и только тогда является языком равномерно ограниченной активной емкости, когда он порождается некоторой грамматикой, надстройка которой является автоматным языком.*

Эта теорема в несколько более слабой форме доказана также в (2).

Институт математики
Сибирского отделения Академии наук СССР
Новосибирск

Поступило
11 XII 1969

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1 Э. Д. Стоцкий. Научно-техническая информация, сер. 2, 5 (1969). 2 J. Friant, Rapport MA-102, CETADOL, Univ, de Montréal (1968).