СВОЙСТВА АКСИОМАТИЧЕСКИХ ГРАММАТИК

Д.В. Дворцовой

Белгосуниверситет, Центр информационных технологий,

Независимости 4, 220050 Минск, Беларусь

dvorc@bsu.by

Введение. В докладе, являющимся продолжением работы [3], рассматриваются варианты грамматик, не использующих метасимволов для промежуточных результатов.

Дано другое доказательство остновной леммы из [3] и уточнено доказательство существования AK-грамматикв [2]. Дано доказательство аксиоматичности всех линейных грамматик. Предложены возможные расширения класса аксиоматических грамматик.

Лемма о ложной вставке. Основная проблема при построении AK-грамматик — это проблема "ложной вставки". Т.е. правило может быть применено не в том месте, где при классическом выводе стоял рекурсивный нетерминал, и в результате, получено слово, не входящее в язык. В [3] было показано, что существуют методы устранения ложных вставок. Была доказана теорема, что в случае регулярных языков, такой процесс будет конечным. Эта теорема опиралась на лемму о подобии цепочек. В формулировке леммы стояло "достаточно большое число". Доказывается теорема, что для достаточно длинной цепочки, это число может быть ограничено до 2-х.

Лемма 1. *Для любых цепочек* a *и* b, $|a| > |b|$, *если найдутся такое* $n \geqslant 2$ *и строка* y, *что* $|y| < |b|$ *и* $a^2 = b^l y$ *то существует такие числа* m, n *и строка* q, *что* $a = q^n$, $b = q^m$.

Доказательство. Найдем $l = \text{НОД}(|a|, |b|)$ и считая $n = |a|/l, m = |b|/l$, разделим строки a и b на подстрочки из l символов.

Получим $a = (a_i)_{i=1..n}$, $b = (b_i)_{i=1..m}$, причем $a_i = b_{i \bmod m}$, откуда $a_i = a_{(i+m) \bmod n}$. Применив формулу к себе, получим $a_i = a_{(i+km) \bmod n}$. Так как m и n взамно просты, существует такое k, что $km \bmod n = 1$, тогда $a_i = a_{i+1}$. Итак, все $a_i = b_i = q$ (кстати, y тоже равно q_1^k).

Формулировка данной леммы отличается от приведенной в [2] для простоты доказательства.

Разбор. Для того, чтобы данные грамматики имели практическое применение, необходимо определить алгоритм разбора и найти классы языков, в которых время разбора не больше, чем для стандартных случаев.

В общем случае разбор (т.е. доказательство того, что данное слово принадлежит языку) невозможен, так как эквивалиентен проблеме соотвествий Маркова. Выделим частные случаи, в которых проблема разбора разрешима.

Определение 1. Разрешимые грамматики.

Неукорачивающая AK-грамматика — та, в которой для каждого правила $(a \to b)$ $|a| \leqslant |b|$.
Монотонная AK-грамматика — та, в которой для каждого правила $(a \to b)$ $|a| < |b|$.

В случае монотонных грамматик разбор можно производить несколькими способами: например, генерировать все слова, пока длина генерируемых слов не более чем длина искомого слова. Если такого слова не найдено, то наше слово не принадлежит языку. Также очевидно, что этот метод требует экспотенциального времени. Для монотонной AK-грамматики данный алгоритм существенно проще.

Возможен и обратный подход — начать с данного слова и кончить какой-либо из аксиом данной грамматики. При этом правила применяются "в обратном направлении".

Определение 2. *Сведение* — операция, при которой, если существует $(x \to y) \in R$ из $a = d_1yd_2$ получается $b = d_1xd_2$. Слово a *непосредственно сводимо* к слову b, если a непосредстенно выводимо из b. Слово a *сводимо* к слову b, если a выводимо из b.

При этом могут получаться и слова, не входящие в язык. Такое слово становится "тупиком" разбора. Применение обращенного правила к такому слову не может дать слово из языка, так как в этом случае мы могли бы получить "тупиковое" слово из аксиомы, применив прямое правило. Наличие тупиков означает, что разбор, возможно, придется проводить с возвратом. Определим подкласс AK-грамматик, в котором разбор линеен.

Определение 3. *Однозначная* AK-грамматика — та, в которой из слова, не принадлежащего языку, невозможно вывести слово, ему принаджлежащее.

Очевидно, разбор однозначной AK-грамматики не будет встречать "тупиков", и достаточно держать в памяти всего одну строку. Все сказанное выше о параллельной обработке тем более справедливо и здесь.

Вопрос в том, для каких классов языков возможно построение однозначных AK-грамматик. Докажем здесь, что для всех регулярных языков существует однозначная AK-грамматика.

Теорема 1. *Грамматика, определенная на основе [3], является однозначной.*

Доказательство. Предположим, существует правило $x \to xy$ и строка $vxyw \in L$ в то время, как $vxw \notin L$. Тогда строка xy была образована в результате применения правила $z \to zt, rzts \in L, zt = pxyq$. То, что $rzt^ns \in L$ и $vxyy^nw \in L$ означает, что согласно лемме из [3] означает, что $t = a^k$.

Остается вопрос неоднозначных грамматик. В [1],гл.2 доказано, что есть существенно неоднозначные языки. Также, не существует алгоритма, позволяющего определить однозначность языка. Из этого следует, что методы, предложенные в [3], не обязательно приведут к желаемому результату.

Время разбора неоднозначных грамматик. Рассмотрим язык

$$G_10 = \{a^{p_1n+p_2m}b^{p_3n+p_4m} \mid n, m \in N\}$$

при условии, что $p_1p_4 - p_2p_3 \neq 0$. AK-грамматика для него очевидна. Для любых допустимых цепочек a^xb^y решение системы уравнений $x = p_1n+p_2m, y = p_3n+p_4m$ будет единственным. Хотя всякий алгоритм анализа, основанный на свертке, должен будет работать с возвратом, время разбора будет линейным, так как для случая $p_1 = 1, p_2 = 2, p_3 = 2, p_4 = 1$ и $a^{3k}b^{3k}$ будут рассмотрены варианты $(n, m) = (0,0), (1,0), \dots, (3k, 0), (0,1), (1,1), \dots, (,1)..(k, k)$

В /КС существует не более чем $O(n)$ слов длины n При однозначности грамматики во время тупика можно не только возвращаться к предыдущей строке, но и применить к тупиковой строке любое из правил. Если задан градиент, то можно прийти к оптимуму за линейное время.

Проблемой однозначных языков с однозначными грамматиками является также необходимость устранения ЛВС. Докажем, что эта проблема разрешима — с помощью расслоения всегда можно добиться AK-грамматики, в которой ЛВС недопустимы. Это будет означать *линейное* время разбора.

Литература

1. *Ахо А., Ульман Дж.* Теория синтаксического анализа, перевода и компиляции, М., 1978. Т. 1.

2. *Мощенский В.А., Дворцовой Д.В.* Контекстно-свободные и Аксиоматические грамматики // Вестн. БГУ. Сер. 1. Физ. Мат. Мех. 1992. N 3.

3. *Дворцовой Д.В.* Аксиоматические и регулярные грамматики. 1994.