Практическое занятие №4. Метод Ангера и метод Кока-Янгера-Касами

Д.Ю. Чалый

24 декабря 2010 г.

1 Пример 1. Метод Ангера.

Разобрать работу метода Ангера на примере следующей грамматики

$$\begin{array}{lcl} S & \to & EvV \\ E & \to & mE \mid lEr \mid mEV \mid V \\ V & \to & i \mid ilEr \mid \varepsilon \end{array}$$

Пусть дана строка $\alpha = mlirvilr$. Начинаем разбирать работу метода Ангера с решения следующей задачи:

S
$mlirvilr$

Возьмем единственную продукцию для S, $S \to EvV$, и, поскольку ее тело состоит из трех символов, мы должны построить все возможные разбиения строки $mlirvilr$ на три класса. Сделаем два замечания. Первое, так как грамматика содержит ε-продукции, то у нас могут быть пустые классы. Второе, в класс, соответствующий терминальному символу, может входить один и только один символ — он сам (действительно, $v \overset{*}{\Rightarrow} v$ и никак иначе). Таким образом, с учетом замечаний, получим одно корректное разбиение:

S		
E	v	V
$mlir$	v	ilr

Итак, мы свели исходную задачу к трем задачам меньшего размера. Теперь нам надо убедиться, что $E \overset{*}{\Rightarrow} mlir$, $v \overset{*}{\Rightarrow} v$ и $V \overset{*}{\Rightarrow} ilr$. Будем решать их последовательно. Рассмотрим первую задачу, $E \overset{*}{\Rightarrow} mlir$ и построим (естественно, с учетом сделанных ранее замечаний), разбиения для первой продукции для E – $E \to mE$:

E	
m	E
m	lir

Таким образом, для того, чтобы убедиться, что $E \overset{*}{\Rightarrow} mlir$, нам надо убедиться, что $m \overset{*}{\Rightarrow} m$ и $E \overset{*}{\Rightarrow} lir$. Первое очевидно, разберемся со вторым. Берем первую продукцию для E и строим все возможные разбиения:

E	
m	E
ε	lir
l	ir
li	r
...	

Очевидно, что не одно из разбиений не подойдет, так как в столбце, помеченном символом m, будут находиться строки, отличные от него. Следовательно, надо взять следующую продукцию для E, продукцию $E \to lEr$. Нам подойдет только одно разбиение:

E		
l	E	r
l	i	r

Таким образом, теперь надо убедиться, что $E \overset{*}{\Rightarrow} i$ (то, что $l \overset{*}{\Rightarrow} l$ и $r \overset{*}{\Rightarrow} r$, очевидно).

Будем последовательно брать продукции для E. Очевидно, что первые три продукции не подойдут, т.к. они содержат терминальные символы, которых нет в строке i. Таким образом необходимо рассмотреть последнюю продукцию, $E \to V$. Получим:

E
V
i

Таким образом, задачу $E \overset{*}{\Rightarrow} i$ мы свели к задаче $V \overset{*}{\Rightarrow} i$. При этом взяв первую продукцию для V мы убедимся что действительно выполняется $V \overset{*}{\Rightarrow} i$. Следовательно, $E \overset{*}{\Rightarrow} i$ и следовательно, $E \overset{*}{\Rightarrow} lir$, из чего следует $E \overset{*}{\Rightarrow} mlir$. Теперь повторяем все рассуждения, но уже для задачи $V \overset{*}{\Rightarrow} ilr$ и убеждаемся, что и эта задача решается. Таким образом, исходная строка выводится в данной грамматике. Обратите внимание, что при описании метода Ангера нельзя упускать его ключевые особенности: построение разбиений и сведение одних задач к другим.

2 Алгоритм Кока-Янгера-Касами

Рассмотрим работу алгоритма Кока-Янгера-Касами (почти) для той же самой грамматики, что и в предыдущем пункте:

$$\begin{array}{lcl} S & \to & EvK \\ E & \to & mE \mid lEr \mid mEK \mid K \\ K & \to & i \mid ilEr \mid \varepsilon \end{array}$$

Алгоритм применим только для грамматик в нормальной форме Хомского. Для преобразования в нормальную форму Хомского необходимо последовательно проводим следующие алгоритмы:

1. Удаляем ε-продукции.
2. Удаляем цепные продукции.

3. Удаляем непорождающие символы.

4. Удаляем недостижимые символы.

5. Приводим к нормальной форме Хомского.

Проделайте это самостоятельно. После первых четырех шагов должна получиться грамматика:

$$\begin{array}{lcl} S & \to & EvK \mid Ev \mid vK \\ E & \to & mE \mid lEr \mid mEK \mid m \mid lr \mid mK \mid i \mid ilEr \mid ilr \\ K & \to & i \mid ilEr \mid ilr \end{array}$$

Приведем эту грамматику в НФХ следующим образом:

$$\begin{array}{rcl} S & \to & ES_{VK} \mid EV \mid VK \\ E & \to & ME \mid LE_{Er} \mid ME_{EK} \mid m \mid LR \mid MK \mid i \mid IE_{lEr} \mid IE_{lr} \\ K & \to & i \mid IE_{lEr} \mid IE_{lr} \\ S_{VK} & \to & VK \\ E_{Er} & \to & ER \\ E_{EK} & \to & EK \\ E_{lEr} & \to & LE_{Er} \\ E_{lr} & \to & LP \\ K & \to & i \mid IE_{lEr} \mid IE_{lr} \\ M & \to & m \\ L & \to & l \\ R & \to & r \\ V & \to & v \\ I & \to & i \end{array}$$

Пусть входной строчкой будет строка $\alpha = mlirvlir$. Алгоритм работает с треугольной таблицей, количество столбцов и строк которой равно длине входной строки:

8								
7								
6								
5								
4								
3								
2								
1								
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

Таблица заполняется снизу вверх. Номер строки таблицы задает длину подстрок исходной строки, которые мы рассматриваем. При заполнении первой строки таблицы мы решаем задачу, из каких нетерминалов выводятся все возможные подстроки длины 1 исходной строки. Очевидно, они могут выводиться только с использованием продукций вида $A \to a$, где $A \in N$, $a \in T$. Таким образом получаем:

8								
7								
6								
5								
4								
3								
2								
1	$\{M, E\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$	$\{V\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

Теперь нам необходимо рассмотреть все подстроки длины 2 исходной строки α. Их будет 7 штук. Первой подстроке ml соответствует ячейка $(1, 2)$. Префикс данной подстроки, m выводится из нетерминалов, расположенных в ячейке $(1, 1)$, а суффикс l – из нетерминалов, расположенных в ячейке $(2, 1)$. Соединим нетерминалы, из которых выводится префикс с нетерминалами, из которых выводится суффикс, получим множество $\{ML, EL\}$. Посмотрим на исходную грамматику, есть ли там продукции с такими телами. Таких продукций нет, следовательно, в ячейку $(1, 2)$ записывается пустое множество:

8								
7								
6								
5								
4								
3								
2	$\varnothing$							
1	$\{M, E\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$	$\{V\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

По аналогии рассматриваем вторую подстроку, li, которой соответствует ячейка $(2, 2)$. Префикс этой подстроки выводится из нетерминалов, расположенных в ячейке $(2, 1)$, а суффикс – из нетерминалов, расположенных в ячейке $(3, 1)$. В итоге мы получим множество из трех элементов $\{LI, LE, LK\}$. Ни один элемент этого множества не является телом какой-либо продукции, следовательно в ячейку $(2, 2)$ тоже будет записано пустое множество. Рассматривая ячейку $(3, 2)$ мы получим множество пар $\{IR, ER, KR\}$. У нас есть продукция $E_{ER} \to ER$. Следовательно подстрока ir выводится из нетерминала E_{ER}. Поместим его в ячейку $(3, 2)$. В итоге заполняя вторую строку таблицы получим:

8								
7								
6								
5								
4								
3								
2	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	
1	$\{M,E\}$	$\{L\}$	$\{I,E,K\}$	$\{R\}$	$\{V\}$	$\{L\}$	$\{I,E,K\}$	$\{R\}$
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

Теперь мы рассматриваем все подстроки длины 3, которым соответствует третья строка таблицы. Рассмотрим подстроку lir, которой соответствует ячейка $(2,3)$. Эта подстрока может быть образована двумя способами. Первый способ — префикс l+суффикс ir. Этому способу соответствует пара ячеек $(2,1)$ (из каких нетерминалов выводится l) и $(3,2)$ (из каких нетерминалов выводится ir). Составим множество пар — $\{LE_{ER}\}$. У нас есть продукция $E_{lEr} \to LE_{ER}$, следовательно, должны поместить нетерминал E_{lEr} в ячейку $(2,3)$. Второй способ — префикс li+суффикс r. Этому способу соответствует пара ячеек $(2,2)$ (из каких нетерминалов выводится li) и $(4,1)$ (из каких нетерминалов выводится r). В данном случае множество пар будет пустым, так как в ячейке $(2,2)$ находится пустое множество. Мы рассмотрели все возможные способы образования подстроки lir и получили, что в ячейку $(2,3)$ необходимо поместить $\{E_{lEr}\}$. Аналогичные рассуждения для остальных ячеек приводят нас к следующей таблице:

8								
7								
6								
5								
4								
3	$\varnothing$	$\{E_{lEr}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{lEr}\}$		
2	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	
1	$\{M,E\}$	$\{L\}$	$\{I,E,K\}$	$\{R\}$	$\{V\}$	$\{L\}$	$\{I,E,K\}$	$\{R\}$
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

Теперь рассмотрим все подстроки длины 4, которым соответствует четвертая строка таблицы. Подстроки длины 4 могут быть образованы четырьмя способами. Необходимо рассмотреть ихвсе чтобы заполнить ячейку. Например, подстрока $lirv$, которой соответствует ячейка $(2,4)$, может быть образована как префикс l+суффикс irv (ячейки $(2,1)$ и $(3,3)$), префикс li+суффикс rv (ячейки $(2,2)$ и $(4,2)$) и префикс lir+суффикс v (ячейки $(2,3)$ и $(5,1)$). Для каждой пары ячеек надо найти множество пар и удостовериться, что у нас есть соответствующие продукции в грамматике. Если такие продукции имеются, то их головы необходимо поместить в ячейку $(2,4)$. Действуя таким образом получим следующую таблицу:

8								
7								
6								
5								
4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$			
3	$\varnothing$	$\{E_{lEr}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{lEr}\}$		
2	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	
1	$\{M, E\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$	$\{V\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

Далее по аналогии рассматриваем остальные строчки таблицы и в итоге получим:

8	$\varnothing$							
7	$\varnothing$	$\varnothing$						
6	$\varnothing$	$\varnothing$	$\varnothing$					
5	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$				
4	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$	$\varnothing$			
3	$\varnothing$	$\{E_{lEr}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{lEr}\}$		
2	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	$\varnothing$	$\varnothing$	$\varnothing$	$\{E_{ER}\}$	
1	$\{M, E\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$	$\{V\}$	$\{L\}$	$\{I, E, K\}$	$\{R\}$
	1	2	3	4	5	6	7	8
	m	l	i	r	v	l	i	r

Так как в ячейке $(1, 8)$ нет начального символа грамматики, то исходная строка не выводится.

3 Задача

При помощи алгоритма Кока-Янгера-Касами убедиться в выводимости строки $mlirvilr$.