УДК 681.142.2

МЕТОД СИНТАКСИЧЕСКИХ КАРТ ДЛЯ ПРОВЕРКИ ВВОДИМЫХ В СИСТЕМУ МАШИН ПРОГРАММ ЗАПИСАННЫХ НА АЛГОРИТМИЧЕСКОМ ЯЗЫКЕ

И.В. Вельбицки, Г.А. Михайлов, Е.Л. Ющенко
(Киев)

В работе [I] описана форма записи грамматик алгоритмических языков в виде синтаксических карт (СКФ), удобная для осуществления их синтаксического анализа.

Пусть $S = (a_1, \ldots, a_n)$ – набор основных символов языка.

Грамматика языка L в СКФ задается словарем термов S и упорядоченным набором синтаксических карт C, каждой из которых соответствует определенный оператор F.

Под синтаксическими картами понимаются булевские матрицы $C^m = \| c^m_{ij} \|$, где m – номер типа синтаксической карты; i – номер строки ($1 \le i \le n$, i-й строке синтаксической карты соответствует символ a_i); j – номер столбца, $1 \le j \le k$ для $m > 1$, $1 \le j \le 2k$ для $m = 1$; k – некоторая константа.

Элементами матриц C^m являются булевские выражения, аргументами которых являются компоненты некоторого Ψ-вектора ($\Psi = \{\psi_1, \ldots, \psi_k\}$) или булевские константы true, false, изображаемые в картах соответственно крестом и пустым символом.

Различаются следующие шесть типов синтаксических карт ($m = I, \ldots, 6$);

C^1 – карта ОШИБОК или карта СОШ;

C^2 – карта ВЗВОДА;

C^3 – карта СБРОСА;

C^4 – карта ФОРМИРОВАНИЯ;

C^5 – карта ЗАПИСИ в СТЭК;

C^6 – карта ЧТЕНИЯ из СТЭКА.

(В примерах для наглядности булевские выражения, не умещающиеся в клетках карт, вынесены за их пределы и обозначены соответствующими номерами).

Действие каждого типа карт задается посредством соответствующего ему оператора F_m , воздействующего на символы проверяемого предложения. Соответственно числу типов синтаксических карт различаются шесть типов F -операторов.

I. F_1 – оператор ПРОВЕРКИ СОВМЕСТИМОСТИ текущего символа a_i с состоянием Ψ -вектора к моменту анализа этого символа:

$$\text{СОШ} := ((\bigvee_{j=1}^{k} \phi_j c^1_{ij}) \vee (\bigvee_{j=1}^{k} \bar{\phi}_j c^1_{i,j+k})),$$

где c^1_{ij} – значение булевского выражения, расположенного на пересечении i -строки и j -столбца карты первого типа;

СОШ – булевская переменная, которой присваивается значение true , если слово с текущим символом в данной позиции не принадлежит языку.

Заметим, что вместо карты "СОШ" (C^1) может быть взята карта "не-СОШ" (условно обозначим её через C^7), оператором действия которой является оператор F_7 :

$$\text{СОШ} := ((\bigwedge_{j=1}^{k} (\bar{\phi}_j \vee c^7_{ij})) \wedge (\bigwedge_{j=1}^{k} (\phi_j \vee c^7_{i,j+k}))).$$

2. F_2 -оператор ВЗВОДА, присваивающий значение true соответствующим компонентам Ψ -вектора:

$$\Psi := \Psi \vee C^2_i .$$

3. F_3 -оператор СБРОСА, присваивающий значение false соответствующим компонентам Ψ -вектора:

$$\Psi := \Psi \wedge \overline{C^3_i} .$$

4. F_4 – оператор ФОРМИРОВАНИЯ, изменяющий на обратные соответствующие значения компонент Ψ -вектора:

$$\Psi := \Psi \oplus C^4_i .$$

Здесь C^m_i ($m = 2, 3, 4$) – булевские вектора, значения компонент которых совпадают с значениями булевских выражений в i -строках соответствующих карт.

5. F_5 -оператор ЗАПИСИ в СТЭК значений компонент Ψ -вектора:

$$\underline{\text{if}} \bigvee_{j=1}^{k} C_{ij}^{5} = \text{true} \ \underline{\text{then}} \quad C_m := \Psi .$$

Запись в стэковую память значений компонент Ψ -вектора осуществляется при просмотре символа типа "открывающей скобки".

6. F_6 - оператор ЧТЕНИЯ из СТЭКА значений компонент Ψ -вектора (ранее записанных в стэковую память оператором F_5):

$$\underline{\text{if}} \bigvee_{j=1}^{k} C_{ij}^{6} = \text{true} \ \underline{\text{then}} \quad \Psi := \Psi \vee C_m .$$

Здесь C_{ij}^{5} и C_{ij}^{6} - значения булевских выражений, расположенных на пересечении i -ой строки и j-го столбца карт C^5 и C^6 . Чтение из стэковой памяти производится при просмотре символа типа "закрывающей скобки".

При проверке предложения в языке используется набор синтаксических карт C , булевская переменная СОШ , стэковая память C_m и вектор Ψ . В начале работы алгоритма $\phi_1 = \ldots = \phi_k =$ СОШ $=$ false стэковая память свободна. Действие алгоритма сводится к посимвольному просмотру заданного предложения слева направо и в применении к каждому символу соответствующих строк операторов синтаксических карт (в принятом в наборе порядке). Схема проверки сводится к следующему:

1) выбирается текущий (вначале первый) символ проверяемого предложения языка;

2) на полученный символ действует оператор F_1 , который проверяет его совместимость текущему состоянию Ψ -вектора . При присвоении переменной СОШ значения true дальнейшая проверка прекращается, иначе:

3) на проверяемый символ последовательно действуют остальные операторы. Проверка предложения заканчивается, если при его просмотре от начала до конца переменная СОШ ни разу не принимала значения true , не происходила выборка из пустого стэка, а стэк в конце проверки пуст. В этом случае предложение считается написанным синтаксически правильно. Предложение не принадлежит языку (синтаксически неправильно), если нарушено хотя бы одно из названных условий.

Введем следующую характеристику $\mathfrak{M}$ набора синтаксических карт C . Пусть $\mathfrak{L}$ - число карт в наборе; $\mathfrak{N}$ - коли-

чество типов карт ($\mathfrak{N} \leq \mathfrak{L}$), последовательность номеров типов – r . Тогда $\mathfrak{M}=[\mathfrak{N}_r, \mathfrak{L}]$ называется характеристикой набора синтаксических карт. В списке r для простоты разделительный знак между номерами типов будем опускать, так как m < 9 . Вторым индексом вверху в обозначении карты будем указывать её порядковый номер в наборе. Наборы С синтаксических карт, могут содержать не все типы синтаксических карт . Например, $\mathfrak{M}=[3_{1434}, 4]$ характеристика класса С карт, состоящего из четырех синтаксических карт трех типов, расположенных в наборе в следующем порядке: C^{11}, C^{42}, C^{33}, C^{44}.

Будем наборы карт, в которых на первом месте находится синтаксическая карта C^1 (если она имеется в наборе), называть нормализованными.

Множество наборов синтаксических карт, имеющих одинаковые характеристики, образует класс синтаксических карт. Ниже будет приведена содержательная характеристика некоторых классов синтаксических карт.

I. Класс $\mathfrak{M}=[2_{14}, 2]$. Языки, определяемые данным классом грамматик, не должны содержать термов типа скобок. Примером такого языка является язык, задаваемый в модифицированной Бэкусово-Науровской форме (МБНФ) записи (см. например,[2]) в виде:

$$1 ::= 2\{a\ 1\} \mid c$$

$$2 ::= ef\ \{d\ k\}$$

Здесь малыми латинскими буквами обозначены терминальные, цифрами-нетерминальные символы (метапеременные) языка; {} – итерационные скобки.

2. Эквивалентными ему в смысле описываемого ими класса языков являются классы с характеристиками $\mathfrak{M}=[3_{123}, 3]$, $\mathfrak{M}=[3_{143}, 3]$.

3. Класс $\mathfrak{M}=[4_{1564}, 4]$, минимальный по числу синтаксических карт, определяющий грамматики, содержащие термы типа скобок. Эквивалентными этому классу являются классы с характеристиками: $\mathfrak{M}=[4_{1654}, 4]$, $\mathfrak{M}=[4_{1645}, 4]$.

4. Класс $\mathfrak{M}=[4_{14564}, 5]$. Данный класс синтаксических карт определяет грамматики большинства известных языков.

Примером языка, принадлежащего данному классу, является

язык, грамматика которого в МБНФ имеет вид.

1 :: = 2{ 4 2}≥| a 2{ 4 2 }*

2 :: = 3| g{g} |3 b 1 {d 1} c|e 1 f (I)

3 :: = k{{ k}{ e }}

4 :: = a|p

Набор синтаксических карт для этого языка приведен на рис.I.

Данному классу эквивалентны классы с характеристиками

$\mathfrak{M}=[4_{14654},5]$, $\mathfrak{M}=[5_{1235623},7]$, $\mathfrak{M}=[6_{123654},6]$.

Остановимся на анализе вида синтаксических карт в некоторых из этих классов.

I. Прежде всего очевидно, что для любого из названных классов пустой язык задается, например, картами, у которых в каждой строке карты C^1 в столбцах от $k+1$ до $2k$ имеется по крайней мере по одному кресту. Очевидно так же, что универсальный язык задается, например, картами, заполненными одними пустыми символами.

2. Если в классе $\mathfrak{M}=[3_{123}, 3]$ наборы синтаксических карт не содержат булевских выражений в качестве своих элементов, то язык, характеризуемый данным классом, является "языком пар символов".Примером такого языка является язык с набором синтаксических карт, изображенным на рисунке 2.В МБНФ этой грамматике соответствует набор правил вида:

1 :: = 2 c

2 :: = a b{a b}

3. Если в классе $\mathfrak{M}=[2_{14}, 2]$ наборы синтаксических карт в качестве своих элементов не содержит булевских переменных, то язык, характеризуемый данным классом, является конечным языком пар символов. Этот язык уже языка, описываемого предшествующим классом синтаксических карт. В МБНФ такому языку соответствует язык, не содержащий рекурсивных определений.

C^{11}

	g	k	l	f	a	b	z	$\bar{g}$	$\bar{k}$	$\bar{l}$	$\bar{f}$	$\bar{a}$	$\bar{b}$	$\bar{z}$
g				×			×							
k	×			×			×							
l	×	×	×	×			×							
f								1		×				
a					×		×							
p								1						
b									×					
c								1					×	
d								1					×	
z			×			×		1						

1 $\bar{l}\,\bar{k}$

C^{42}

	g	k	l	f	a	b	z
g							
k							
l					1		
f							
a							
p							
b					2		
c							
d							
z							

1 $\overline{\bar{l}\,\bar{a}}\,\bar{b} \vee l\,a \vee b\,a$

2 $\bar{l}\,\bar{b}$

C^{53}

	g	k	l	f	a	b	z
g							
k							
l			×				
f							
a							
p							
b						×	
c							
d							
z							

C^{64}

	g	k	l	f	a	b	z
g							
k							
l							
f				×			
a							
p							
b							
c						×	
d							
z							

C^{45}

	g	k	l	f	a	b	z
g	$\bar{g}$				×		
k		$\bar{k}$			×		
l			$\bar{l}$		a	b	
f	g	k	2	$\bar{f}$	a		
a	g	k	f		×		
p	g	k	f		×		
b		×	l			$\bar{b}$	
c	g	k		$\bar{f}$	a	1	
d	g	k	f				
z	×	×		×			×

1 $a \vee l$

2 $a \vee l$

Рис. I.

Примером такого языка является язык

$$1 ::= a\ 2$$

$$2 ::= b\ 3$$

$$3 ::= c$$

который в СКФ описывается набором карт,изображенным на рис.3.

Сказанное не означает, что класс грамматик с характеристикой $\mathfrak{M} = [2_{14}, 2]$ слабее класса $\mathfrak{M} = [3_{123}, 3]$. Очевидно, что обе эти грамматики описывают один и тот же класс языков. Только при этом вид карт первого из названных классов проще чем второго, хотя число карт во второй грамматике меньше, чем в первой. Например, язык, описанный картами, приведенными на рис.2, в классе $\mathfrak{M} = [2_{14}, 2]$ описывается синтаксическими картами, изображенными на рис.3.

Заметим, что в языках, используемых на практике, часто встречаются термы, играющие в языке, с синтаксической точки зрения, одну и ту же роль. Таковы, например,в языке Алгол- 60 символы $\leq$, $>$, $=$, $\neq$, $<$, $\geq$ или знаки + , - и т.д. Наличие групп таких символов в языке позволяет для целей контроля упростить запись их грамматики в СКФ.

Определение. Два терминальных символа a , b в языке L назовем синтаксически эквивалентными, если для любых терминальных цепочек ω , φ языка из $\omega a \varphi \in L$ следует $\omega b \varphi \in L$, и наоборот.

Очевидно, что терминальные символы грамматики согласно данного определения распадаются на классы взаимно эквивалентных символов. Выберем по одному представителю из каждого класса и назовем их синтермами (синтаксическими термами) языка.

Преследуя цели контроля языков, вместо карт, задаваемых в терминальном словаре, можно рассматривать карты в синтерминальном словаре, что зачастую приводит к существенному сокращению объема карт.

Например, легко видеть, что для языка, задаваемого в МБНФ в виде:

$$1 ::= 2\{4\ 2\} \mid 6\ 2\{4\ 2\}$$

$$2 ::= 3 \mid g\{g\} \mid 3\ b\ 1\ \{d\ 1\}\ c \mid e\ 1\ f$$

C^{11}

	a	б	c	$\bar{a}$	$\bar{б}$	$\bar{c}$
a	X		X			
б		X	X	X		
c	X		X		X	

C^{22}

	a	б	c
a	X		
б		X	
c			X

C^{33}

	a	б	c
a		X	X
б	X		X
c	X	X	

Рис. 2.

C^{11}

	a	б	c	$\bar{a}$	$\bar{б}$	$\bar{c}$
a	X		X			
б		X	X	X		
c	X		X		X	

C^{42}

	a	б	c
a	X	б	
б	X	X	
c		X	X

Рис. 3.

$$
\begin{aligned}
3 &::= 7\{\{7\}\{g\}\} \\
4 &::= 6 \mid 5 \\
5 &::= p \mid q \mid t \\
6 &::= a \mid y \\
7 &::= k \mid m \mid n \mid r
\end{aligned}
\qquad (2)
$$

имеет место следующая таблица соответствия

Термы	g	k	m	n	r	e	f	a	y	p	q	t	b	c	d	z
Синтермы	g	k	k	k	k	e	f	a	a	p	p	p	b	c	d	z

,

учет которой позволяет перейти к записи грамматики с СКФ, совпадающей с грамматикой (I). В этом случае синтаксическому контролю по грамматике (2) с помощью синтаксических карт должен предшествовать переход от термов к синтермам.

В заключение отметим, что описанный метод синтаксических карт может быть реализован как программно, так и схемно в виде простого автономного от машины устройства синтаксического контроля [2]. При программной реализации метода универсальная программа синтаксического контроля языков, грамматики которых описаны в СКФ, занимает около 50 команд машины типа М-20 и для практических языков (например, автокод "Инженер" машины Минск [3]) работает со скоростью в среднем 30 логических команд на один символ проверяемого предложения языка.

При схемной реализации метода устройство синтаксического контроля занимает менее одного стандартного шкафа универсальной вычислительной машины. Время работы устройства составляет менее 1% времени работы существующих устройств ввода и полностью с ним совмещено.

ЛИТЕРАТУРА

1. И.В. Вельбицкий, Г.А. Михайлов, Е.Л. Ющенко. Метод синтаксических карт для проверки программ, журн. "Кибернетика", 1967, № 3.

2. И.В. Вельбицкий, Г.А. Михайлов, Е.Л. Ющенко. Об одном методе синтаксического контроля программ, записанных на алгоритмическом языке. Труды семинара "Алгоритмические языки и автоматизация программирования", вып. I, 1967.

3. М.Е. Неменман, В.И. Цегельский, И.М. Матюшевская. Автокод для решения инженерных задач на машине "Минск-2", Минск, 1965.