

Лекція 13-14.

**ГРАМАТИКИ ДЛЯ РОЗПІЗНАВАННЯ
ОБРАЗІВ**

В случае неоднозначности, т. е. когда некоторая цепочка имеет два или несколько грамматических разборов, в качестве синтаксического описания этой цепочки можно взять, например, наиболее вероятный вывод, т. е. вывод, порождающий данную цепочку с наибольшей вероятностью. Если цепочки $x \in L(\bar{G}_s)$, представляющие образы, и связанные с этими цепочками вероятности $p(x)$ известны (или их можно оценить), то можно так вычислить или оценить вероятности правил подстановки, чтобы полученная стохастическая грамматика G_s порождала данные цепочки приблизительно с исходными вероятностями. Кроме того, приписывание порождаемым цепочкам вероятностей может служить средством отсева «нежелательных» цепочек. Этим цепочкам можно приписать малые исходные вероятности так, чтобы получаемая в результате стохастическая грамматика с большой вероятностью порождала только желательные цепочки. Рассмотрим ряд наглядных примеров.

Пример. 6.1. На рис. 6.1 показан равносторонний треугольник и восемь его искаженных вариантов, а на рис. 6.2 — выбранные производные элементы. Каждый треугольник описывается цепочкой длины 3. Предположим, что на основе априорных знаний или текущих исследований известны или могут быть оценены вероятности этих девяти треугольников. Например, эти вероятности можно оценить по относительным частотам появления фигур.

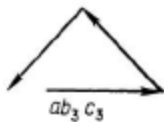
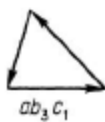


Рис. 6.1. Треугольники для примера 6.1.

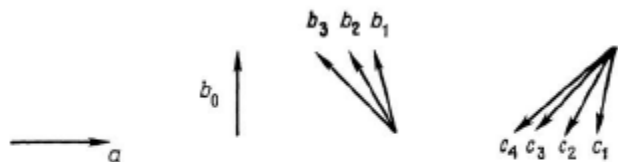


Рис. 6.2. Непроизводные элементы для примеров 6.1 и 6.2.

Эта информация приведена в табл. 6.1.

Таблица 6.1

x	$p(x)$	x	$p(x)$
$x_1 = ab_1c_1$	1/36	$x_6 = ab_2c_3$	2/36
$x_2 = ab_1c_2$	2/36	$x_7 = ab_3c_1$	3/36
$x_3 = ab_1c_3$	3/36	$x_8 = ab_3c_2$	2/36
$x_4 = ab_2c_1$	1/36	$x_9 = ab_3c_3$	1/36
$x_5 = ab_2c_2$	21/36		

Все указанные цепочки порождает стохастическая регулярная грамматика ¹⁾ $G_1 = (V_{N1}, V_{T1}, P_1, S)$, где $V_{N1} = \{S, A_1, A_2,$

$A_3, A_4\}$, $V_{T1} = \{a, b_1, b_2, b_3, c_1, c_2, c_3\}$ и P_1 :

$$1) S \xrightarrow{p_1} aA_1, \quad 8) A_3 \xrightarrow{p_8} c_1,$$

$$2) A_1 \xrightarrow{p_2} b_1A_2, \quad 9) A_3 \xrightarrow{p_9} c_2,$$

$$3) A_1 \xrightarrow{p_3} b_2A_3, \quad 10) A_3 \xrightarrow{p_{10}} c_3,$$

$$4) A_1 \xrightarrow{p_4} b_3A_4, \quad 11) A_4 \xrightarrow{p_{11}} c_1,$$

$$5) A_2 \xrightarrow{p_5} c_1, \quad 12) A_4 \xrightarrow{p_{12}} c_2,$$

$$6) A_2 \xrightarrow{p_6} c_2, \quad 13) A_4 \xrightarrow{p_{13}} c_3.$$

$$7) A_2 \xrightarrow{p_7} c_3,$$

Из условий нормировки следуют уравнения

$$\begin{aligned} p_1 = 1, \quad p_2 + p_3 + p_4 = 1, \quad p_5 + p_6 + p_7 = 1, \\ p_8 + p_9 + p_{10} = 1, \quad p_{11} + p_{12} + p_{13} = 1. \end{aligned}$$

Для порождения цепочки $x_1 = ab_1c_1$ нужны правила подстановки 1), 2) и 5), следовательно,

$$p(x_1) = p_1 p_2 p_5 = 1/36.$$

Аналогично,

$$\begin{aligned} p(x_2) = p_1 p_2 p_6 = 2/36, & \quad p(x_6) = p_1 p_3 p_{10} = 2/36, \\ p(x_3) = p_1 p_2 p_7 = 3/36, & \quad p(x_7) = p_1 p_4 p_{11} = 3/36, \\ p(x_4) = p_1 p_3 p_8 = 1/36, & \quad p(x_8) = p_1 p_4 p_{12} = 2/36, \\ p(x_5) = p_1 p_3 p_9 = 21/36, & \quad p(x_9) = p_1 p_4 p_{13} = 1/36. \end{aligned}$$

Из уравнений $p_1 = 1$, $p_1 p_2 (p_5 + p_6 + p_7) = 1/36 + 2/36 + 3/36 = 1/6$, $p_5 + p_6 + p_7 = 1$ получим $p_2 = 1/6$, $p_5 = 1/6$, $p_6 = 1/3$, $p_7 = 1/2$. Аналогично,

$$p_3 = 2/3, \quad p_8 = 1/24, \quad p_9 = 21/24, \quad p_{10} = 1/12,$$

$$p_4 = 1/6, \quad p_{11} = 1/2, \quad p_{12} = 1/3, \quad p_{13} = 1/6.$$

Грамматика G_1 с вычисленными вероятностями правил подстановки порождает цепочки в точности с теми же вероятностями, которые приведены в табл. 6.1.

Пример 6.2. На рис. 6.3. показан прямоугольный треугольник и восемь его искаженных вариантов. Если использовать непрое-

водные элементы, показанные на рис. 6.2, а вероятности взять из табл. 6.2, то девять цепочек — описаний этих фигур — поро-

ждает стохастическая регулярная грамматика $G_2 = (V_{N_2}, V_{T_2}, P_2, S)$, где

$$V_{N_2} = \{S, A_1, A_2, A_3, A_4\}, \quad V_{T_2} = \{a, b_0, b_1, b_2, c_2, c_3, c_4\}$$

и P_2 :

$$\begin{array}{lll} S \xrightarrow{1} aA_1, & A_2 \xrightarrow{1/12} c_3, & A_3 \xrightarrow{1/6} c_4, \\ A_1 \xrightarrow{2/3} b_0A_2, & A_2 \xrightarrow{21/24} c_4, & A_4 \xrightarrow{1/31} c_2, \\ A_1 \xrightarrow{1/6} b_1A_3, & A_3 \xrightarrow{1/6} c_2, & A_4 \xrightarrow{1/2} c_3, \\ A_1 \xrightarrow{1/6} b_2A_4, & A_3 \xrightarrow{2/3} c_3, & A_4 \xrightarrow{1/6} c_4, \\ A_2 \xrightarrow{1/24} c_2. & & \end{array}$$

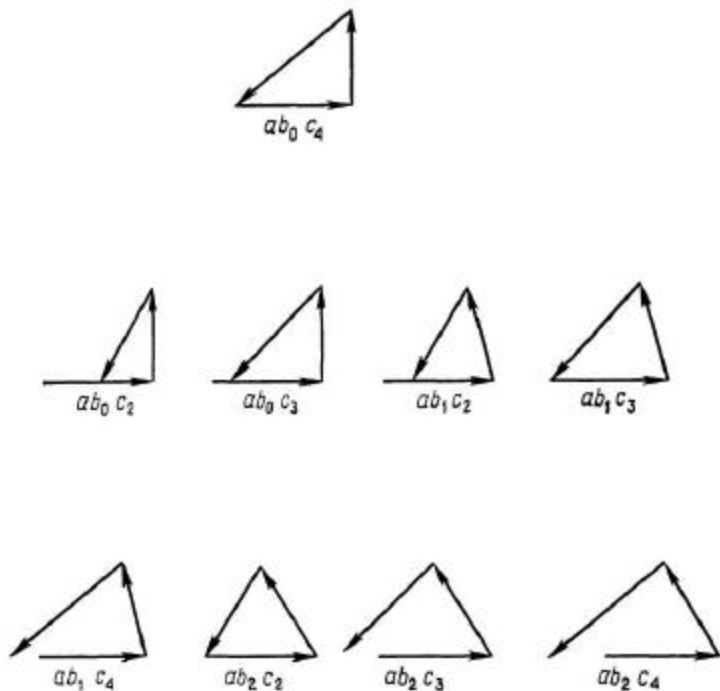


Рис. 6.3. Треугольники для примера 6.2.

Заметим, что цепочки ab_1c_2 , ab_1c_3 , ab_2c_2 и ab_2c_3 из табл. 6.2 содержатся и в табл. 6.1. Будем считать, что объекты, описываемые цепочками из табл. 6.1, составляют класс ω_1 , а объекты, описываемые цепочками из табл. 6.2 — класс ω_2 . Тогда четыре объек-

та ab_1c_2 , ab_1c_3 , ab_2c_2 и ab_2c_3 попадут сразу в оба класса. В детерминистском случае конечные автоматы (алгоритмы синтаксического анализа), синтезированные на основе грамматик \bar{G}_1 и \bar{G}_2 , не смогут различить принадлежности этих четырех цепочек. Иными словами, эти цепочки будут допущены обоими автоматами. В подобных ситуациях для определения принадлежности объектов можно использовать идею дискриминантного подхода. В задаче распознавания существенную роль играет информация об условных вероятностях $p(x/\omega_i)$, $i = 1, 2$, цепочек, принадлежащих к каждому классу. Например, предположим, что распознаваемый объект представлен цепочкой ab_1c_2 . Предположим, что априорные вероятности классов равны (т. е. $P(\omega_1) = P(\omega_2)$). В данном случае $p(ab_1c_2 | G_1) = 2/36$, а $p(ab_1c_2 | G_2) = 1/36$. По правилу максимального правдоподобия объект, представленный цепочкой ab_1c_2 , следует отнести к классу ω_1 , поскольку $p(ab_1c_2 | G_1) > p(ab_1c_2 | G_2)$.

Таблица 6.2

x	$p(x)$	x	$p(x)$
ab_0c_2	$1/36$	ab_1c_4	$1/36$
ab_0c_3	$2/36$	ab_2c_2	$2/36$
ab_0c_4	$21/36$	ab_2c_3	$3/36$
ab_1c_2	$1/36$	ab_2c_4	$1/36$
ab_1c_3	$4/36$		

Поскольку в правиле классификации используются вероятности $p(x | G_i)$, их вычисление для каждой поступающей цепочки x становится важной задачей. В совсем простых случаях, как в примере 6.2, эту информацию в виде табл. 6.1 и 6.2 (которую легко вывести из двух соответствующих стохастических грамматик) можно непосредственно хранить в памяти распознающей системы. Необходимые для классификации данные можно находить при помощи процедуры «просмотр таблицы». Однако в общем случае число порождаемых грамматикой цепочек часто очень велико или бесконечно. Использование конечного представления вместо выписывания всех цепочек является как раз одним из преимуществ грамматики в описании структуры языка. В случае регулярных языков для распознавания можно синтезировать конечные автоматы. Например, при помощи описанной в разд. 5.3.1 процедуры можно построить стохастический конечный автомат S_{a1} , который воспринимает цепочки, порождаемые грамматикой G_1 из примера 6.1.

Этот автомат определен пятеркой

$$S_{a1} = (\Sigma, Q, M, \pi_0, F),$$

где

$$\Sigma = \{a, b_1, b_2, b_3, c_1, c_2, c_3\}, \quad Q = \{S, A_1, A_2, A_3, A_4, T, R\}, \\ F = \{T\}, \quad \pi_0 = [1, 0, 0, 0, 0, 0, 0]$$

и

$$M(a) = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad M(b_1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & \frac{1}{6} & 0 & 0 & 0 & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

$$M(b_2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad M(b_3) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M(c_1) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad M(c_2) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M(c_3) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{12} & \frac{11}{12} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & \frac{5}{6} \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \pi_F = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Легко проверить, что

$$p(ab_1c_2 | G_1) = \pi_0 M(ab_1c_2) \pi_F = \pi_0 M(a) M(b_1) M(c_2) \pi_F = 2/36.$$

Аналогично,

$$p(ab_1c_1 | G_1) = \pi_0 M(ab_1c_1) \pi_F = 1/36,$$

$$p(ab_1c_3 | G_1) = \pi_0 M(ab_1c_3) \pi_F = 3/36,$$

$$p(ab_2c_1 | G_1) = \pi_0 M(ab_2c_1) \pi_F = 1/36,$$

$$p(ab_2c_2 | G_1) = \pi_0 M(ab_2c_2) \pi_F = 21/36,$$

$$p(ab_2c_3 | G_1) = \pi_0 M(ab_2c_3) \pi_F = 2/36,$$

$$p(ab_3c_1 | G_1) = \pi_0 M(ab_3c_1) \pi_F = 3/36,$$

$$p(ab_3c_2 | G_1) = \pi_0 M(ab_3c_2) \pi_F = 2/36,$$

$$p(ab_3c_3 | G_1) = \pi_0 M(ab_3c_3) \pi_F = 1/36.$$

Для образов, описываемых бесконтекстными языками, требуется синтаксический анализ исходной цепочки (образа) по отношению к каждой грамматике G_i . После нахождения всех грамматических разборов можно вычислить вероятности $p(x | G_i)$ порождения цепочки x каждой грамматикой. Если цепочка x не может быть порождена грамматикой G_j , то считают, что

$p(x | G_j) = 0$. Затем применяют правило классификации по максимальному правдоподобию, т. е. относят представленный цепочкой x объект к классу ω_k , если

$$p(x | G_k) = \text{Max}_{G_i} \{p(x | G_i)\}. \quad (6.1)$$

Грамматический разбор по отношению к грамматике G_k дает синтаксическое описание цепочки x . На рис. 6.4 показана блок-схема такого распознающего устройства. В более общем случае можно сформулировать правила классификации, учитывая известные априорные вероятности и (или) функции потерь, например решающее правило Байеса.

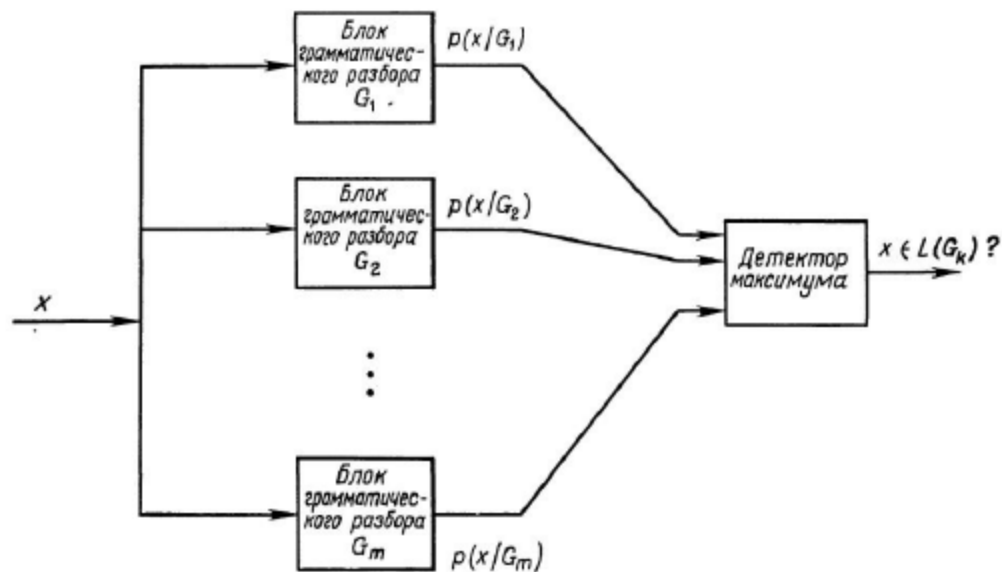


Рис. 6.4. Блок-схема стохастического узнающего устройства, работающего по принципу максимума правдоподобия.

Пример 6.3. Рассмотрим грамматику, которая предназначена для описания «зашумленных» квадратов. Такие фигуры можно получать, например, на экране катодно-лучевой трубки вводного устройства или на табло фирмы RAND [4]. Благодаря некоторым ограничениям эта грамматика довольно проста, но интересна. Ограничения таковы.

1) Длина стороны квадрата является случайной величиной с геометрическим распределением и со средним значением $1/\alpha$ ($0 < \alpha < 1$).

2) Стороны квадрата ориентированы по горизонтали и вертикали. Предполагается, что оператор «рисует» квадрат настолько аккуратно, что «шум» обусловлен в основном ошибкой квантования в аналого-цифровом преобразователе.

3) Эффект «улучшения качества действия при наблюдении цели», который способствует замыканию фигуры, когда она почти замкнута, не учитывается.

Эти упрощающие ограничения можно отбросить и сделать грамматику произвольно сложной. Однако данный пример в первую очередь рассчитан на демонстрацию следующих положений:

1) бесконтекстная программная грамматика (БПГ) может порождать относительно сложный язык непосредственно составляющих;

2) при помощи стохастической БПГ можно моделировать физический процесс со свойственными ему шумами;

3) общая вычислительная программа, моделирующая действие программных грамматик, является полезным средством разработки грамматик в режиме диалога.

Программа моделирует цепочки и выводит на дисплей некоторую графическую интерпретацию процесса порождения. Возможность порождения цепочек и вывода на дисплей позволяет наблюдать действие грамматики, улучшать ее и качественно оценивать ее стохастические свойства.

В табл. 6.3 приведена стохастическая БПГ, а на рис. 6.5 — типичный пример «зашумленного» квадрата. Вероятности пере-

ходов были выбраны, исходя из «разумных соображений». При наличии уже проанализированного в рамках данной грамматической модели «обучающего множества» эти вероятности можно

было бы подобрать в лучшем соответствии с физическим процессом.

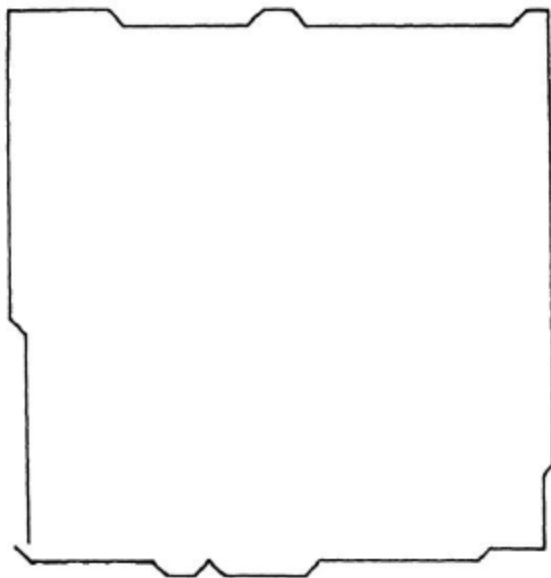


Рис. 6.5. Типичный «зашумленный квадрат». Масштаб $\sim 1,6$ мм на клетку
растра; $\alpha = 0,05$, $\beta = 0,05$.

Стохастическая грамматика для описания «зашумленных» квадратов

$$G_{sp} = (V_N, V_T, P, S, J),$$

где

$$V_N = \{S, F, X, Q, A, M, W, C, E, R, T, B, V\}, \quad V_T = \{-, 0, +, *, \$\},$$

$$J = \{1, 2, \dots, 37\}$$

и

0: отрезок единичной длины в направлении данной стороны

+: отрезок единичной длины с наклоном $+1$ по отношению к данной стороне

—: отрезок единичной длины

*: угол (изменение направления на 90°)

§: ограничитель остатка (только для окончательной цепочки)

<i>Метка</i>	<i>Ядро</i>	<i>Список переходов при успехе и вероятности переходов</i>	<i>Список переходов при неудаче и вероятности переходов</i>
1	$S \rightarrow FXVQ$	$\{2, 4, 7\}, \{1 - 2\alpha, \alpha, \alpha\}$	$\{\emptyset\}, \{1\}$
2	$F \rightarrow QFA$	$\{3\}, \{1\}$	$\{\emptyset\}, \{1\}$
3	$X \rightarrow XXX$	$\{10\}, \{1\}$	$\{\emptyset\}, \{1\}$
4	$F \rightarrow +FA$	$\{5\}, \{1\}$	$\{\emptyset\}, \{1\}$
5	$M \rightarrow XXX$	$\{10\}, \{1\}$	$\{6\}, \{1\}$
6	$X \rightarrow WXX$	$\{10\}, \{1\}$	$\{\emptyset\}, \{1\}$
7	$F \rightarrow -FA$	$\{8\}, \{1\}$	$\{\emptyset\}, \{1\}$
8	$W \rightarrow XXX$	$\{10\}, \{1\}$	$\{9\}, \{1\}$
9	$X \rightarrow MXX$	$\{10\}, \{1\}$	$\{\emptyset\}, \{1\}$
10	$W \rightarrow W$	$\{2, 4, 7, 12\}, \{p_1, p_2, p_3, \beta\}$	$\{11\}, \{1\}$
11	$M \rightarrow M$	$\{2, 4, 7, 12\}, \{p_1, p_3, p_2, \alpha\}$	$\{2, 4, 7, 12\}, \{p_4, p_3, p_3, \beta\}$
12	$F \rightarrow *F$	$\{13\}, \{1\}$	$\{\emptyset\}, \{1\}$
13	$Q \rightarrow R$	$\{16\}, \{1\}$	$\{14\}, \{1\}$
14	$R \rightarrow S$	$\{16\}, \{1\}$	$\{15\}, \{1\}$
15	$S \rightarrow T$	$\{16\}, \{1\}$	$\{36\}, \{1\}$
16	$B \rightarrow A$	$\{16\}, \{1\}$	$\{17\}, \{1\}$
17	$A \rightarrow B$	$\{18\}, \{1\}$	$\{20\}, \{1\}$
18	$E \rightarrow C$	$\{17\}, \{1\}$	$\{19\}, \{1\}$
19	$V \rightarrow CV$	$\{17\}, \{1\}$	$\{\emptyset\}, \{1\}$
20	$W \rightarrow X$	$\{21\}, \{1\}$	$\{22\}, \{1\}$
21	$C \rightarrow E$	$\{20\}, \{1\}$	$\{\emptyset\}, \{1\}$
22	$M \rightarrow X$	$\{23\}, \{1\}$	$\{25, 26, 29\}, \{1 - 2\alpha, \alpha, \alpha\}$
23	$E \rightarrow C$	$\{22\}, \{1\}$	$\{24\}, \{1\}$
24	$V \rightarrow CV$	$\{22\}, \{1\}$	$\{\emptyset\}, \{1\}$
25	$F \rightarrow 0F$	$\{32\}, \{1\}$	$\{\emptyset\}, \{1\}$
26	$F \rightarrow +F$	$\{27\}, \{1\}$	$\{\emptyset\}, \{1\}$
27	$M \rightarrow X$	$\{32\}, \{1\}$	$\{28\}, \{1\}$
28	$X \rightarrow W$	$\{32\}, \{1\}$	$\{\emptyset\}, \{1\}$

29	$F \rightarrow \neg F$	$\{30\}, \{1\}$	$\{\emptyset\}, \{1\}$
30	$W \rightarrow X$	$\{32\}, \{1\}$	$\{31\}, \{1\}$
31	$X \rightarrow M$	$\{32\}, \{1\}$	$\{\emptyset\}, \{1\}$
32	$C \rightarrow E$	$\{33\}, \{1\}$	$\{\emptyset\}, \{1\}$
33	$C \rightarrow C$	$\{34\}, \{1\}$	$\{12\}, \{1\}$
34	$W \rightarrow W$	$\{25, 26, 29\}, \{1 - \alpha - \alpha^2, \alpha^2, \alpha\}$	$\{35\}, \{1\}$
35	$M \rightarrow M$	$\{25, 26, 29\}, \{1 - \alpha - \alpha^2, \alpha, \alpha^2\}$	$\{25, 26, 29\}, \{1 - 2\alpha, \alpha, \alpha\}$
36	$T \rightarrow S$	$\{37\}, \{1\}$	$\{\emptyset\}, \{1\}$
37	$F \rightarrow F$	$\{\emptyset\}, \{1\}$	$\{\emptyset\}, \{1\}$

Квадрат на рис. 6.5 представлен цепочкой—

$$\begin{aligned}
 & -0000000000 - 00 + -000000 + 000000000000 + 0000* \\
 & 00000 - 00* \\
 & 00 + 0000000000000000 - 00 + 000000000 - 0000000* \\
 & 000000000000000000000000 + 0000000000000000* \S \text{ (остаток опущен)}
 \end{aligned}$$

Здесь $p_1 = (1-\alpha-\alpha^2)(1-\beta)$, $p_2 = \alpha^2(1-\beta)$, $p_3 = \alpha(1-\beta)$, $p_4 = (1-2\alpha)(1-\beta)$.

Пример 6.4. Этот пример служит иллюстрацией другого применения стохастических БПГ в описании образов. Следующие три грамматики предназначены для описания изображений равноплечных, неравноплечных и акроцентрических хромосом (см. рис. А.2). Они построены, исходя из эвристических соображений, связанных с тем, чтобы избежать неоднозначности и порождения некоторых нежелательных объектов такого типа, как показано на

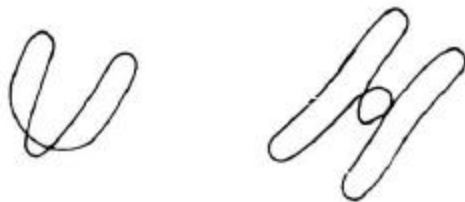


Рис. 6.6. Нежелательные изображения хромосом

рис. 6.6. Данные о применении бесконтекстных грамматик при описании изображений хромосом содержатся в приложении А.

Вероятности переходов в этих трех грамматиках определяют распределения вероятностей изображений хромосом каждого из трех классов. Грамматики построены так, что некоторые неравноплечные хромосомы рассматриваются как искаженные или зашумленные варианты акроцентрических, а равноплечные и акроцентрические хромосомы — как искаженные варианты неравноплечных. Поэтому вероятности ($p_{1m}, \dots, p_{4m}, p_{1s}, \dots, p_{7s}, p_{1a}, \dots, p_{5a}$) выбраны так, чтобы 1) равноплечные хромосомы чаще порождались грамматикой G_m , чем G_s или G_a , 2) неравноплечные чаще порождались грамматикой G_s , чем G_m или G_a и 3) акроцентрические чаще порождались грамматикой G_a , чем G_m или G_s .

Равноплечные хромосомы описываются грамматикой

$$G_m = (V_{Tm}, V_{Nm}, J_m, P_m, S),$$

где

$$V_{Tm} = \{a, b, c, d\},$$

где a, b, c и d — производные элементы, показанные на рис. А.1,

$$V_{Nm} = \{S, A, B, C, D\}, \quad J_m = \{1, \dots, 25\},$$

а P_m определены следующим образом:

Μεταβλητά	Άθροισμα	U	P(U)
1	$S \rightarrow ABC$	2	1
2	$A \rightarrow cA$	3	1
3	$B \rightarrow BcB$	4, 10	$P_{1m}, 1 - P_{1m}$
4	$C \rightarrow Cb$	5	1
5	$A \rightarrow bA$	6	1
6	$B \rightarrow Db$	7	1
7	$B \rightarrow bD$	8	1
8	$D \rightarrow B$	9	1
9	$D \rightarrow B$	4, 10	$P_{2m}, 1 - P_{2m}$
10	$A \rightarrow aA$	11	1
11	$C \rightarrow Ca$	12	1
12	$B \rightarrow Da$	13	1
13	$B \rightarrow aD$	14	1
14	$D \rightarrow B$	15	1
15	$D \rightarrow B$	16, 22	$P_{3m}, 1 - P_{3m}$
16	$A \rightarrow A$	17	1
17	$C \rightarrow Cb$	18	1
18	$B \rightarrow Db$	19	1
19	$B \rightarrow bD$	20	1
20	$D \rightarrow B$	21	1
21	$D \rightarrow B$	22, 16	$P_{4m}, 1 - P_{4m}$
22	$A \rightarrow B$	23	1
23	$C \rightarrow b$	24	1
24	$B \rightarrow db$	25	1
25	$B \rightarrow bd$	\emptyset	1

Неравноплечные хромосомы описываются грамматикой

$$G_S = (V_{TS}, V_{NS}, J_S, P_S, S),$$

где $V_{TS} = V_{Tm}$, $V_{NS} = \{S, A, B\}$, $J_S = \{1, \dots, 20\}$, а P_S определены следующим образом:

Метка	Ядро	U	$P(U)$
1	$S \rightarrow AB$	2	1
2	$A \rightarrow cA$	3, 5	$P_{1S}, 1 - P_{1S}$
3	$A \rightarrow bA$	4	1
4	$B \rightarrow Bb$	3, 5	$P_{2S}, 1 - P_{2S}$
5	$A \rightarrow aA$	6	1
6	$B \rightarrow Ba$	7	1
7	$A \rightarrow bA$	8	1

8	$B \rightarrow Bb$	7, 9	$P_{3s}, 1 - P_{3s}$
9	$A \rightarrow dA$	10	1
10	$A \rightarrow Bd$	11	1
11	$A \rightarrow bA$	12	1
12	$B \rightarrow Bb$	11, 13, 17	$P_{4s}, P_{5s}, 1 - P_{4s} - P_{5s}$
13	$A \rightarrow aA$	14	1
14	$B \rightarrow Ba$	15, 19	$P_{6s}, 1 - P_{6s}$
15	$A \rightarrow bA$	16	1
16	$B \rightarrow Bb$	15, 19	$P_{7s}, 1 - P_{7s}$
17	$A \rightarrow a$	18	1
18	$B \rightarrow ca$	\emptyset	1
19	$A \rightarrow b$	20	1
20	$B \rightarrow cb$	\emptyset	1

Акроцентрические хромосомы описываются грамматикой

$$G_a = (V_{Ta}, V_{Na}, J_a, P_a, S),$$

где $V_{Ta} = V_{Tm}$, $V_{Na} = \{S, A, B\}$, $J_a = \{1, \dots, 18\}$, а P_a определены следующим образом:

Метка	Ядро	U	$P(U)$
1	$S \rightarrow AB$	2	1
2	$A \rightarrow cA$	3	1
3	$A \rightarrow aA$	4	1
4	$B \rightarrow Ba$	5	1
5	$A \rightarrow bA$	6	1
6	$B \rightarrow Bb$	5, 7	$P_{1a}, 1 - P_{1a}$
7	$A \rightarrow dA$	8	1
8	$B \rightarrow Bd$	9	1
9	$A \rightarrow bA$	10	1
10	$B \rightarrow Bb$	9, 11, 15	$P_{2a}, P_{3a}, 1 - P_{2a} - P_{3a}$
11	$A \rightarrow aA$	12	1
12	$B \rightarrow Ba$	13, 17	$P_{4a}, 1 - P_{4a}$
13	$A \rightarrow bA$	14	1
14	$B \rightarrow Bb$	13, 17	$P_{5a}, 1 - P_{5a}$
15	$A \rightarrow a$	16	1
16	$B \rightarrow ca$	\emptyset	1
17	$A \rightarrow b$	18	1
18	$B \rightarrow cb$	\emptyset	1