

УДК 519.2:004.421.5:004.7

В. Н. ЗАДОРЖНЫЙ  
Е. Б. ЮДИН  
М. Н. ЮДИНА

Омский государственный  
технический университет,  
г. Омск

Институт математики  
им. С. Л. Соболева СО РАН,  
Омский филиал

## РАСПРЕДЕЛЕНИЯ СТЕПЕНЕЙ В РАСТУЩИХ ГРАФАХ, ТЕРЯЮЩИХ ДУГИ

Для растущих графов предпочтительного связывания, непрерывно теряющих дуги, решается задача расчета двумерного распределения степеней связности дуг (ребер). Применение разработанных методов расчета графов с потерями дуг позволяет синтезировать адекватные модели растущих сетей (социальных, информационно-телекоммуникационных, сетей сотрудничества и т.д.), учитывающие потери связей между узлами. Тем самым расширяются возможности использования таких сетей и эффективного управления их развитием, в том числе путем влияния на процессы потери связей.

**Ключевые слова:** случайные графы с нелинейным правилом предпочтительного связывания, распределение степеней связности вершин, распределение степеней связности дуг, калибровка случайных графов.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 16-31-60023 мол\_а\_дк.*

**Введение.** Наблюдаемый в последние годы интенсивный расцвет теории случайных графов обусловлен тем, что она является математической основой науки о сетях — нового раздела физики, изучающего объективные законы развития сетей и сетевых процессов, которые не только радикально расширяют возможности современной человеческой цивилизации, но и порождают проблемы, от решения которых зависит само ее существование. Предметом исследования науки о сетях являются информационные, транспортные, социальные и многие другие сети, противоборство в информационных сетях, распространение болезнетворных вирусов и т.д.

Наиболее распространенными моделями сетей являются *растущие графы* предпочтительного связывания. Первая их версия — графы Барабаши — Альберт (графы БА) предложены и исследованы в работах [1–3]. Плодотворность изложенных в этих рабо-

тах идей ознаменовалась последующим взрывным ростом числа публикаций о растущих сетях (см., например, [4–15]). В [10–15] для моделирования сетей предлагаются и исследуются растущие графы с нелинейным правилом предпочтительного связывания (НППС). Граф с НППС [11] выращивается из небольшого графа-затравки путем добавления к нему в моменты  $t = t_1, t_2, \dots$  очередного *приращения графа* — новой вершины с ограниченным случайным числом  $x \leq h$  исходящих из нее дуг. Концы этих дуг связываются со случайно выбираемыми вершинами графа. Вероятность  $p_i$  того, что дуга выберет для связи вершину  $i$ , пропорциональна весу  $f$  этой вершины, определяемому через ее степень связности  $k_i$ :

$$p_i = f(k_i) / \sum_{j=1}^N f(k_j), \quad (1)$$

где  $N$  — текущее число вершин в графе. При неограниченном добавлении приращений формируется бесконечный граф с НППС.

Таким образом, случайный граф с НППС задается двумя параметрами — распределением  $\{r_k\}$  вероятностей случайной величины  $x$  — числа дуг в приращении и весовой функцией  $f(k) \geq 0$ . Будем считать, что если  $k \notin [0, M]$  (где  $M \leq \infty$ ), то  $f(k) = 0$ . Далее функцию  $f(k)$  целочисленной переменной  $k$  будем обозначать в виде  $f_k$ . При линейном весе ( $f_k = k$ ) и фиксированном числе дуг  $x = m = \text{const}$  в приращениях граф с НППС становится графом БА.

Одной из наиболее важных асимптотических характеристик растущего графа является распределение степени связности (РСС)  $\{Q_k\}$  его вершин,  $k=0, 1, 2, \dots$ . Здесь  $Q_k$  — это вероятность того, что случайно (равновероятно) выбранная в бесконечном графе вершина имеет степень связности  $k$ .

Не менее важными асимптотическими характеристиками растущего графа являются РСС его дуг  $\{Q_{l,k}\}$  и РСС ребер  $\{\Theta_{l,k}\}$ . Здесь  $Q_{l,k}$  — это вероятность того, что случайно выбранная дуга (с.в.д.) ориентированного бесконечного графа исходит из вершины со степенью  $l$  и заходит в вершину со степенью  $k$ . Аналогично  $\Theta_{l,k}$  — вероятность того, что случайно выбранное ребро неориентированного бесконечного графа, проходящее в случайно выбранном направлении, ведет из вершины со степенью  $l$  в вершину со степенью  $k$  [10].

Появление работ [16–18] свидетельствует о росте внимания к проблеме потерь в растущих сетях. Найденные в [17, 18] РСС вершин полностью согласуются с результатами, которые мы получили в [13, 14] численными методами для графов с НППС, теряющих вершины или ребра. Однако мы не нашли публикаций, решающих задачу о распределении степеней дуг или ребер в таких графах. Ниже эта задача решается для графов с потерями дуг путем вывода и решения уравнений, определяющих искомые распределения.

**Растущие графы, теряющие дуги.** Непрерывные потери связей в ходе развития сетей моделируются графами с потерями дуг. На каждом шаге выращивания графа, заданного параметрами  $\{r_k\}$  и  $\{f_k\}$ , последовательно выполняются две операции:

1) к графу добавляется по правилу (1) приращение со случайным числом  $x$  дуг, в среднем с  $M(x) = m$  дугами;

2) из графа удаляется случайное число с.в.д., в среднем  $\gamma$  дуг ( $0 < \gamma < m$ ).

Финальные распределения  $\{Q_k\}$ ,  $\{Q_{l,k}\}$  и  $\{\Theta_{l,k}\}$  не зависят от затравки графа. На каждом шаге времени общее число дуг возрастает в среднем на  $\dot{m} = m - \gamma$  и число дуг в растущем графе сходится к  $\dot{m}N$ , где  $N$  — число вершин графа.

Финальное РСС  $\{Q_{l,k}\}$  дуг будем находить в виде матрицы  $\mathbf{Q} = \parallel Q_{l,k} \parallel$ . Неориентированный граф получается из выращиваемого ориентированного графа заменой всех дуг ребрами. Матрица  $\Theta = \parallel \Theta_{l,k} \parallel$  финального РСС ребер вычисляется по формуле  $\Theta = (\mathbf{Q} + \mathbf{Q}^T) / 2$ , где  $T$  — символ транспонирования.

**Распределение степеней вершин.** Для нахождения РСС  $\mathbf{Q}$  и  $\Theta$  воспользуемся найденным в [13] РСС  $\{Q_k\}$  вершин графов с потерями дуг:

$$Q_k = \frac{r_k + \frac{mQ_{k-1}f_{k-1} + k + 1}{\langle f \rangle} \gamma Q_{k+1}}{1 + \frac{mf_k}{\langle f \rangle} + \frac{k\gamma}{\dot{m}}}, \quad 0 \leq k \leq M + 1,$$

$$\langle f \rangle = \sum_{k=0}^M f_k Q_k, \quad (2)$$

где  $\langle f \rangle$  — средний вес вершин графа, и  $Q_{-1} = 0$ . Система уравнений (2) быстро и точно решается в электронных таблицах при выборе режима «итерации», разрешающего циклические ссылки ячеек друг на друга и на самих себя.

Рекомендуется пересчет листа инициировать вручную и начинать итерации при записанной в ячейку для  $\langle f \rangle$  константе 1. Затем, после появления ненулевых  $Q_k$ , следует ввести в ячейку для  $\langle f \rangle$  расчетную формулу (2) и продолжить итерации до получения неизменяющихся результатов.

**Распределения степеней дуг и ребер.** Для вывода уравнений и формул, позволяющих вычислять РСС  $\{Q_{l,k}\}$  дуг, определим слой  $A_k$ ,  $k \geq 0$ , как множество вершин графа, имеющих степень  $k$ . Туннель  $B_{l,k}$  определим как множество дуг, исходящих из вершин слоя  $A_l$  и заходящих в вершины слоя  $A_k$  [10].

Придерживаясь сформированной в [10, 11] схемы вывода уравнений баланса, запишем для  $Q_{l,k}$  асимптотически точное уравнение баланса в виде

$$\dot{m}(N + 1)Q_{l,k} = \dot{m}NQ_{l,k} + \Delta_1 + \Delta_2, \quad l, k \geq 1. \quad (3)$$

где  $\dot{m}(N + 1)Q_{l,k}$  — среднее число дуг в туннеле  $B_{l,k}$  на шаге  $N + 1$ , т.е. после добавления  $(N + 1)$ -й вершины;  $\dot{m}NQ_{l,k}$  — среднее число дуг в туннеле  $B_{l,k}$  на шаге  $N$ ;  $\Delta_1$  — средний прирост числа дуг в туннеле  $B_{l,k}$  за счет нового приращения графа;  $\Delta_2$  — средний прирост числа дуг в туннеле  $B_{l,k}$  (имеющий отрицательное значение) за счет потери в среднем  $\gamma$  с.в.д. графа. Из (3) имеем:

$$\dot{m}Q_{l,k} = \Delta_1 + \Delta_2, \quad l, k \geq 1. \quad (4)$$

Прирост  $\Delta_1$  формируется здесь так же, как в графе без потерь [12]:

$$\Delta_1 = lr_l P_{k-1} + mP_{k-1} \frac{|B_{l,k-1}|}{|A_{k-1}|} + mP_{l-1} \frac{|B_{l-1,k}|}{|A_{l-1}|} - mP_k \frac{|B_{l,k}|}{|A_k|} - mP_l \frac{|B_{l,k}|}{|A_l|}. \quad (5)$$

Через  $P_j$  в (5) обозначена вероятность связывания дуги приращения с вершиной слоя  $A_j$ . В [12] показано, что  $P_j \sim f_j Q_j / \langle f \rangle$ . Прямые скобки в (5) используются для обозначения числа элементов множества. Отношение  $|B_{l,k-1}| / |A_{k-1}|$  сходится к  $\dot{m}NQ_{l,k-1} / (NQ_{k-1}) = \dot{m}Q_{l,k-1} / Q_{k-1}$ . Аналогично  $|B_{l-1,k}| / |A_{l-1}| \sim \dot{m}Q_{l-1,k} / Q_{l-1}$ ,  $|B_{l,k}| / |A_k| \sim \dot{m}Q_{l,k} / Q_k$  и  $|B_{l,k}| / |A_l| \sim \dot{m}Q_{l,k} / Q_l$ . Подставляя в (5) приведенные выражения, включая выражения для  $P_j$ , получаем

$$\Delta_1 = lr_l \frac{f_{k-1} Q_{k-1}}{\langle f \rangle} + m \dot{m} \left( \frac{f_{k-1}}{\langle f \rangle} Q_{l,k-1} + \frac{f_{l-1}}{\langle f \rangle} Q_{l-1,k} - \frac{f_l}{\langle f \rangle} Q_{l,k} - \frac{f_k}{\langle f \rangle} Q_{l,k} \right). \quad (6)$$

Выражение для прироста  $\Delta_2$  в (4) найдем следующим образом. Разобьем множество  $\Omega$  элементарных исходов при выборе теряемой с.в.д.  $\bar{e}$  на непересекающиеся подмножества  $\Omega_i$  (случаи) и определим *эффекты*  $E(\Omega_i)$ , т.е. произведения вероятностей  $P(\Omega_i)$  случаев и соответствующих условных средних приростов числа дуг в  $B_{l,k}$ . Безусловный средний прирост, вызванный потерей одной

Оценки первых двадцати эффектов потери дуги

Эффект	Оценка эффекта	Эффект	Оценка эффекта
$E(\bar{e} \in B_{l,l})$	$-2D_l Q_{l,l}$	$E(\bar{e} \in B_{l,k+1})$	$-D_l Q_{l,k+1}$
$E(\bar{e} \in B_{k,k})$	$-2D_k Q_{k,k}$	$E(\bar{e} \in B_{k+1,l})$	$D_{k+1} Q_{k+1,l} - D_l Q_{k+1,l}$
$E(\bar{e} \in B_{l+1,l+1})$	$2D_{l+1} Q_{l+1,l+1}$	$E(\bar{e} \in B_{k,l+1})$	$-D_k Q_{k,l+1} + D_{l+1} Q_{k,l+1}$
$E(\bar{e} \in B_{k+1,k+1})$	$2D_{k+1} Q_{k+1,k+1}$	$E(\bar{e} \in B_{l+1,k})$	$-D_k Q_{l+1,k}$
$E(\bar{e} \in B_{l,k})$	$-1 \cdot Q_{l,k}$	$E(\bar{e} \in B_{k,k+1})$	$-D_k Q_{k,k+1} + D_{k+1} Q_{k,k+1}$
$E(\bar{e} \in B_{k,l})$	$-D_k Q_{k,l} - D_l Q_{k,l}$	$E(\bar{e} \in B_{k+1,k})$	$D_{k+1} Q_{k+1,k} - D_k Q_{k+1,k}$
$E(\bar{e} \in B_{l,l+1})$	$-D_l Q_{l,l+1} + D_{l+1} Q_{l,l+1}$	$E(\bar{e} \in B_{l+1,k+1})$	$D_{l+1} Q_{l+1,k+1} + D_{k+1} Q_{l+1,k+1}$
$E(\bar{e} \in B_{l+1,l})$	$D_{l+1} Q_{l+1,l} - D_l Q_{l+1,l}$	$E(\bar{e} \in B_{k+1,l+1})$	$D_{k+1} Q_{k+1,l+1} + D_{l+1} Q_{k+1,l+1}$

с.в.д., равен сумме найденных эффектов. Поэтому  $\Delta_2$  найдем, умножая сумму найденных эффектов на среднее число  $\gamma$ , теряемых с.в.д.

Потеря с.в.д.  $\bar{e}$  влияет на число дуг в  $B_{l,k}$  в следующих 20 случаях:

— в четырех случаях, когда выбранная для удаления (теряемая) с.в.д.  $\bar{e}$  соединяет две вершины одного и того же слоя из слоев  $A_l, A_k, A_{l+1}, A_{k+1}$ ,

— в  $6 \times 2 = 12$  случаях, когда теряемая с.в.д.  $\bar{e}$  соединяет в том или ином направлении два слоя из слоев  $A_l, A_k, A_{l+1}, A_{k+1}$ ,

— в четырех случаях, когда с.в.д.  $\bar{e}$  инцидентна вершине одного из четырех перечисленных слоев и вершине, не принадлежащей ни одному из них.

Во всех остальных случаях потеря с.в.д. не изменяет числа дуг в туннеле  $B_{l,k}$ .

При определении эффектов  $E(\Omega_i)$  пригодятся следующие асимптотически точные выражения для среднего числа дуг того или иного туннеля, приходящихся на одну вершину того или иного слоя:

$$\begin{aligned}
 D_l &= \frac{|B_{l,k}|}{|A_l|} \sim \frac{\dot{m}NQ_{l,k}}{NQ_l} = \frac{\dot{m}Q_{l,k}}{Q_l}, \\
 D_k &= \frac{|B_{l,k}|}{|A_k|} \sim \frac{\dot{m}Q_{l,k}}{Q_k}, \\
 D_{l+1} &= \frac{|B_{l+1,k}|}{|A_{l+1}|} \sim \frac{\dot{m}Q_{l+1,k}}{Q_{l+1}}, \\
 D_{k+1} &= \frac{|B_{l,k+1}|}{|A_{k+1}|} \sim \frac{\dot{m}Q_{l,k+1}}{Q_{k+1}}. \tag{7}
 \end{aligned}$$

Оценки для первых 16 подмножеств исходов, влияющих на число дуг в туннеле  $B_{l,k}$ , приведены в табл. 1. Способ получения этих оценок поясним на примере оценки эффекта  $E(\bar{e} \in B_{l,l+1})$  (табл. 1).

Если теряемая с.в.д.  $\bar{e}$  принадлежит туннелю  $B_{l,l+1}$ , то в результате ее потери происходят следующие изменения. Степень вершины  $v_1 \in A_l$ , из которой исходит дуга  $\bar{e}$ , в результате потери этой дуги уменьшается на единицу. Поэтому вершина переходит из слоя  $A_l$  в слой  $A_{l-1}$ , и инцидентные ей дуги туннеля  $B_{l,k}$  (в среднем приблизительно  $D_l$  дуг) переходят в туннель  $B_{l-1,k}$ . Из-за этого число дуг в туннеле  $B_{l,k}$  уменьшается в среднем приблизительно на  $D_l$ . Вместе с тем вершина  $v_2 \in A_{l+1}$ , в которую заходит дуга  $\bar{e}$ , переходит из-за потери этой дуги в слой  $A_l$ ,

что добавляет в состав туннеля  $B_{l,k}$  в среднем приблизительно  $D_{l+1}$  инцидентных вершине  $v_2$  дуг туннеля  $B_{l+1,k}$ . Упомянутые здесь величины  $D_l$  и  $D_{l+1}$  приведены в (7). Общее изменение  $(D_{l+1} - D_l)$  числа дуг, умноженное на вероятность  $Q_{l,l+1}$  рассматриваемого случая  $\bar{e} \in B_{l,l+1}$ , составляет величину эффекта  $E(\bar{e} \in B_{l,l+1}) = -D_l Q_{l,l+1} + D_{l+1} Q_{l,l+1}$ . Аналогично выводятся оценки всех первых двадцати эффектов (табл. 1).

Для оставшихся четырех случаев необходимо определить вероятность  $P[\bar{e} \in I(A_j)]$  принадлежности с.в.д.  $\bar{e}$  множеству  $I(A_j)$  дуг, инцидентных вершинам слоя  $A_j$  ( $j = l, k, l+1, k+1$ ). Все вершины слоя  $A_j$  имеют степень  $j$ , и некоторые дуги в  $I(A_j)$  инцидентны сразу двум вершинам в  $A_j$ . Поэтому  $P[\bar{e} \in I(A_j)] = (j | A_j | - | B_{j,j} |) / (\dot{m}N) \sim (jNQ_j - \dot{m}NQ_{j,j}) / (\dot{m}N) = jQ_j / \dot{m} - Q_{j,j}$ . Обозначая объединение непересекающихся множеств плюсом, вычитание подмножества — обратной наклонной, находим последние четыре эффекта:

$$\begin{aligned}
 E(\bar{e} \in I(A_l) \setminus (B_{l,l} + B_{l,k} + B_{l,l+1} + \\
 + B_{l,k+1} + B_{k,l} + B_{l+1,l} + B_{k+1,l})) \approx \\
 \approx -D_l \cdot \left( \frac{lQ_l}{\dot{m}} - Q_{l,l} - Q_{l,l} - Q_{l,k} - Q_{l,l+1} - \right. \\
 \left. - Q_{l,k+1} - Q_{k,l} - Q_{l+1,l} - Q_{k+1,l} \right), \\
 E(\bar{e} \in I(A_k) \setminus (B_{k,k} + B_{k,l} + B_{k,l+1} + \\
 + B_{k,k+1} + B_{l,k} + B_{l+1,k} + B_{k+1,k})) \approx \\
 \approx -D_k \cdot \left( \frac{kQ_k}{\dot{m}} - Q_{k,k} - Q_{k,k} - Q_{k,l} - Q_{k,l+1} - \right. \\
 \left. - Q_{k,k+1} - Q_{l,k} - Q_{l+1,k} - Q_{k+1,k} \right), \\
 E(\bar{e} \in I(A_{l+1}) \setminus (B_{l+1,l+1} + B_{l+1,l} + B_{l+1,k+1} + B_{l+1,k} + B_{l,l+1} + \\
 + B_{k+1,l+1} + B_{k+1,l})) \approx D_{l+1} \times \\
 \times \left( \frac{l+1}{\dot{m}} Q_{l+1} - Q_{l+1,l+1} - Q_{l+1,l+1} - Q_{l+1,l} - Q_{l+1,k+1} - \right. \\
 \left. Q_{l+1,k} - Q_{l,l+1} - Q_{k+1,l+1} - Q_{k,l+1} \right), \\
 E(\bar{e} \in I(A_{k+1}) \setminus (B_{k+1,k+1} + B_{k+1,k} + B_{k+1,l} + B_{k+1,l+1} +
 \end{aligned}$$

I10		fx = ЕСЛИ(\$D\$2=0;0;(\$F10*\$E10*\$H\$5*\$H\$6/\$B\$4/\$B\$5+\$B\$3*\$H10*\$H\$6/\$B\$5+\$B\$3*\$I9*\$D															
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	
1				Пуск													
2	гамма	1		1													
3	m	3															
4	m²	2															
5	<f>	1															
6								Qk:	0.01238	0.02476	0.11148	0.32222	0.21915	0.1392	0.08238	0.04541	0.023
7								fk:	0	1	1	1	1	1	1	1	1
8								k:	0	1	2	3	4	5	6	7	8
9	0	0.01238		0	0	0		0	0	0	0	0	0	0	0	0	0
10	1	0.02476		1	0	1		0	0.0002	0.00088	0.00203	0.0031	0.00338	0.003	0.00229	0.001	
11	2	0.11148		1	0	2		0	0.00085	0.0044	0.01154	0.01949	0.0205	0.01715	0.01233	0.007	
12	3	0.32222		1	1	3		0	0.00187	0.01328	0.04175	0.08741	0.08207	0.0817	0.04083	0.024	
13	4	0.21915		1	0	4		0	0.00189	0.00893	0.02434	0.04396	0.04849	0.03822	0.02889	0.016	
14	5	0.1392		1	0	5		0	0.00117	0.00535	0.0133	0.0222	0.02478	0.02155	0.01586	0.01	
15	6	0.08238		1	0	6		0	0.0007	0.00291	0.00683	0.01094	0.01251	0.01128	0.00857	0.00	
16	7	0.04541		1	0	7		0	0.00037	0.00146	0.0033	0.00517	0.00598	0.00552	0.00429	0.002	
17	8	0.02333		1	0	8		0	0.00018	0.00068	0.0015	0.00232	0.00271	0.00254	0.002	0.001	
18	9	0.01119		1	0	9		0	8E-05	0.0003	0.00064	0.00099	0.00116	0.0011	0.00087	0.000	
19	10	0.00501		1	0	10		0	3.3E-05	0.00012	0.00026	0.0004	0.00047	0.00045	0.00038	0.000	
20	11	0.0021		1	0	11		0	1.3E-05	4.7E-05	9.9E-05	0.00015	0.00018	0.00017	0.00014	9.8E-	
21	12	0.00082		1	0	12		0	4.7E-06	1.7E-05	3.6E-05	5.4E-05	6.4E-05	6.2E-05	5E-05	3.6E-	
22	13	0.00029		1	0	13		0	1.6E-06	5.5E-06	1.2E-05	1.8E-05	2.1E-05	2E-05	1.7E-05	1.2E-	
23	14	8.3E-05		2	0	14		0	4.2E-07	1.5E-06	3.1E-06	4.7E-06	5.6E-06	5.5E-06	4.5E-06	3.2E-	
24	15	3.9E-05		3	0	15		0	1.9E-07	6.6E-07	1.4E-06	2.1E-06	2.5E-06	2.4E-06	2E-06	1.4E-	
25	16	2.3F-05		4	0	16		0	1.1F-07	3.7F-07	7.8F-07	1.2F-06	1.4F-06	1.4F-06	1.1F-06	8F-	

Рис. 1. Расчет РСС дуг в тестовом графе с НППС, теряющем дуги

$$+ B_{k,k+1} + B_{l,k+1} + B_{l+1,k+1})) \approx D_{k+1} \times$$

$$\times \left( \frac{k+1}{m} Q_{k+1} - Q_{k+1,k+1} - Q_{k+1,k+1} - Q_{k+1,k} - Q_{k+1,l} - \right.$$

$$\left. - Q_{k+1,l+1} - Q_{k,k+1} - Q_{l,k+1} - Q_{l+1,k+1} \right).$$

Суммируя эти эффекты с приведенными в табл. 1, раскрывая скобки, приводя подобные члены и подставляя выражения (7), получаем средний прирост числа дуг в туннеле  $B_{l,k}$ , обусловленный потерей одной с.в.д. Умножая его на  $\gamma$ , получаем искомый средний прирост  $\Delta_2$  за шаг выращивания графа

$$\Delta_2 \approx \gamma(l+1)Q_{l+1,k} + \gamma(k+1)Q_{l,k+1} - \gamma(l+k+1)Q_{l,k} +$$

$$+ \gamma \frac{\dot{m}Q_{l,k}^2}{Q_l} + \gamma \frac{\dot{m}Q_{l,k}^2}{Q_k} - \gamma \frac{\dot{m}Q_{l+1,k}^2}{Q_{l+1}} - \gamma \frac{\dot{m}Q_{l,k+1}^2}{Q_{k+1}}. \quad (8)$$

Подставляя теперь выражения  $\Delta_2$  (8) и  $\Delta_1$  (6) в (4), получаем относительно искомых вероятностей  $Q_{l,k}$  приближенную систему уравнений

$$\dot{m}Q_{l,k} = l r_l \frac{f_{k-1} Q_{k-1}}{\langle f \rangle} + m \dot{m} \frac{f_{k-1}}{\langle f \rangle} Q_{l,k-1} + m \dot{m} \frac{f_{l-1}}{\langle f \rangle} Q_{l-1,k} -$$

$$- m \dot{m} \frac{f_l}{\langle f \rangle} Q_{l,k} - m \dot{m} \frac{f_k}{\langle f \rangle} Q_{l,k} +$$

$$+ \gamma(l+1)Q_{l+1,k} + \gamma(k+1)Q_{l,k+1} - \gamma \frac{\dot{m}Q_{l+1,k}^2}{Q_{l+1}} - \gamma \frac{\dot{m}Q_{l,k+1}^2}{Q_{k+1}} -$$

$$- \gamma(l+k+1)Q_{l,k} + \gamma \frac{\dot{m}Q_{l,k}^2}{Q_l} + \gamma \frac{\dot{m}Q_{l,k}^2}{Q_k}, \quad (l, k \geq 1), \quad (9)$$

из которой выводим формулу расчета  $Q_{l,k}$  методом простых итераций:

$$Q_{l,k} =$$

$$= \frac{l r_l \frac{f_{k-1} Q_{k-1}}{\langle f \rangle} + m \frac{f_{k-1}}{\langle f \rangle} Q_{l,k-1} + m \frac{f_{l-1}}{\langle f \rangle} Q_{l-1,k} +}{1 + m \frac{f_l}{\langle f \rangle} + m \frac{f_k}{\langle f \rangle} + \gamma} \rightarrow$$

$$\rightarrow \gamma \left( \frac{(l+1)}{\dot{m}} Q_{l+1,k} + \frac{(k+1)}{\dot{m}} Q_{l,k+1} - \frac{Q_{l+1,k}^2}{Q_{l+1}} - \frac{Q_{l,k+1}^2}{Q_{k+1}} \right)$$

$$\rightarrow \left( \frac{(l+k+1)}{\dot{m}} - \frac{Q_{l,k}}{Q_l} - \frac{Q_{l,k}}{Q_k} \right) \quad (10)$$

( $l, k \geq 1$ ). Решение (9), (10) задачи о РСС дуг является приближенным, поскольку при его выводе использованы приближенные оценки эффектов.

**Эксперименты.** В Excel для расчета  $Q_{l,k}$  формулу (10) рекомендуется ввести в ячейку, соответствующую элементу  $Q_{l,1}$  матрицы  $Q$  (рис. 1), правильно закрепляя в формуле ссылки на параметры графа и на ячейки верхних и левых «заголовков» матрицы, рассчитанных заранее. Так, в примере, показанном на рис.1, введенную в ячейку I10 формулу (10) в виде

$$= \text{ЕСЛИ}(\$D\$2=0;0;(\$F10*\$E10*\$H\$5*\$H\$6/\$B\$4/\$B\$5 + \$B\$3*\$H10*\$H\$6/\$B\$5 +$$

$$\$B\$3*\$I9*\$D9/\$B\$5 + \$B\$2*((\$F10 + 1)*I11/\$B\$4 - I11^2/\$B11 + (I\$7 + 1)*J10/\$B\$4 - J10^2/J\$5)) /$$

$$(1 + \$B\$3*I\$6/\$B\$5 + \$B\$3*\$D10/\$B\$5 + \$B\$2*(\$F10 + I\$7 + 1)/\$B\$4 - I10/\$B10 - I10/I\$5))$$

можно просто скопировать из ячейки I10 сразу на весь прямоугольный диапазон ячеек рассчитываемой части матрицы  $Q$ .

Тестовый граф, расчет распределения  $\{Q_{l,k}\}$  которого показан на рис. 1, определяется параметрами  $f_k = 1$  при  $k=0, \dots, 13$  и  $f_k = k - 12$  при  $k > 13$ ;  $r_3 = 1$  (т.е.  $m = 3$ );  $\gamma = 2$ ;  $\dot{m} = m - \gamma = 2$ . Распределение  $\{Q_k\}$  и средний вес  $\langle f \rangle$  для этого графа вычислены заранее путем решения системы уравнений (2). Ячейка D2 (Пуск) используется для «сбрасывания в ноль» результатов расчета матрицы  $Q = \|Q_{l,k}\|$ . Нулевая строка (для  $l=0$ ) и нулевой столбец (для  $k=0$ ) матрицы  $Q$  заполняются нулями (константами). Левый и верхний «заголовки» продлеваются на одну ячейку дальше границ заполненного формулами диапазона. Итерации при расчете листа сходятся за несколько секунд. Параллельно с матрицей  $Q$  на этом же листе рассчитываются транспонированная матрица  $Q^T$ , матрица РСС ребер и строится график РСС ребер (рис. 2).

На рис. 2 рассчитанное РСС ребер тестового графа сравнивается с распределением, полученным путем имитационного моделирования (ИМ) графа, т.е. путем его непосредственного выращивания. Сравнение показывает, что в данном случае погрешности формулы (10) невелики. Хорошую точность формулы (10) подтверждают и ее проверки посредством расчета и ИМ ряда других графов с потерями дуг.

Точность формулы (10) и сходимость порождаемых ею итераций ухудшаются с ростом доли теряемых дуг. Тем не менее, пример с тестовым графом,

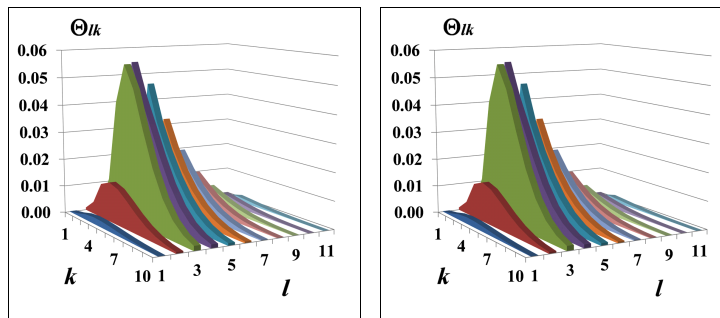


Рис. 2. РСС ребер тестового графа: расчетное (слева) и полученное путем ИМ (справа)

теряющим 33 % дуг (см. рис. 2), показывает, что для исследования многих реальных сетей точность формулы (10) достаточна. К их числу относится и сеть веб-страниц, поскольку она теряет лишь 15 % дуг [19].

**Заключение.** В статье разработан численный метод расчета двумерного распределения степеней дуг и двумерного распределения степеней ребер в растущих графах, теряющих дуги (ребра).

Хотя разработанный численный метод является приближенным, он имеет точность, достаточную для исследования многих реальных растущих сетей с потерями связей.

Полученные результаты позволяют калибровать графовые модели реальных сетей с потерями связей одновременно по распределениям степеней вершин и распределениям степеней дуг (ребер) подобно тому, как это делается при моделировании сетей без потерь [18]. Тем самым обеспечивается возможность существенного повышения адекватности графовых моделей, синтезируемых для повышения эффективности использования социальных, информационных и других сетей и для разработки стратегий воздействия на их развитие.

#### Библиографический список

1. Barabasi A. L., Albert R. Emergence of scaling in random networks // *Science*. 1999. Vol. 286. P. 509–512. DOI: 10.1126/science.286.5439.509.
2. Barabasi A. L., Albert R. Statistical mechanics of complex networks // *Rev. Mod. Phys.* 2002. Vol. 74. P. 47–97.
3. Barabasi A. L. Scale-free networks: A decade and beyond // *Science*. 2009. Vol. 325. P. 412–413. DOI: 10.1126/science.1173299.
4. Krapivsky P. L., Redner S. Organization of growing random networks // *Phys. Rev. E*. 2001. Vol. 63. P. 066123. DOI: 10.1103/PhysRevE.63.066123.
5. Newman M. The structure of scientific collaboration networks // *Proc. Natl. Acad. Sci. USA*. 2001. Vol. 98. P. 404–409. DOI: 10.1007/978-88-470-0665-2\_13.
6. Antal T., Krapivsky P., Redner S. Dynamics of social balance on networks // *Physical Review E*. 2005. Vol. 72 (3). P. 036121. DOI: 10.1103/PhysRevE.72.036121.
7. Clauset A., Shalizi C. R., Newman M. Power-law distributions in empirical data // *Rev. Mod. Phys.* 2009. Vol. 81. P. 661–703. DOI: 10.1137/070710111.
8. Cohen R., Havlin S. *Complex networks: structure, stability and function* // Cambridge University Press. 2010. DOI: 10.1017/S09780511780356.
9. Ghoshal G., Chi L., Barabasi A. L. Uncovering the role of elementary processes in network evolution // *Scientific Reports*. 2013. Vol. 3. P. 1–8. DOI: 10.1038/srep02920.

10. Zadorozhnyi V. N., Yudin E. B. Structural properties of the scale-free Barabasi-Albert graph // *Automation and Remote Control*. 2012. Vol. 73. № 4. P. 702–716.

11. Zadorozhnyi V. N., Yudin E. B. Growing network: models following nonlinear preferential attachment rule // *Physica A: Statistical Mechanics and its Applications*. 2015. Vol. 428. P. 111–132. DOI: 10.1016/j.physa.2015.01.052.

12. Задорожный В. Н. Растущие сети: динамика распределения степеней связности смежных узлов // *Омский научный вестник*. 2016. № 2 (146). С. 81–86.

13. Задорожный В. Н., Юдин Е. Б. Уравнения динамики степеней узлов в растущих сетях с потерями связей // *Динамика систем, механизмов и машин*. 2016. № 1, Т. 3. С. 340–346.

14. Задорожный В. Н. Растущие сети с потерями узлов // *Омский научный вестник*. 2017. № 151. С. 108–113.

15. Zadorozhnyi V. N., Yudin E. B., Yudina M. N. Analytical and numerical methods of calibration for preferential attachment random graphs // 2017 International Siberian Conference on Control and Communications (SIBCON), Astana, Kazakhstan. 2017. P. 1–6. DOI: 10.1109/SIBCON.2017.7998461.

16. West R., Paranjape A., Leskovec J. Mining Missing Hyperlinks from Human Navigation Traces: A Case Study of Wikipedia // *Proceedings of the 24th International Conference on World Wide Web*. 2015. P. 1242–1252.

17. Barabasi, A. L. *Network Science* // Cambridge University Press. 2015. URL: <http://barabasi.com/networksciencebook> (дата обращения: 25.09.2017).

18. Ghoshal G., Chi L., Barabasi A.-L. Uncovering the role of elementary processes in network evolution // *Scientific Reports*. 2013. Vol. 3. P. 1–8.

19. Fenner T., Levene M., Loizou G. A stochastic model for the evolution of the web allowing link deletion // *ACM Transactions in Internet Technology*. 2006. Vol. 6, № 2. P. 117–130.

**ЗАДОРЖНЫЙ Владимир Николаевич**, доктор технических наук, доцент (Россия), профессор кафедры «Автоматизированные системы обработки информации и управления» Омского государственного технического университета (ОмГТУ).

Адрес для переписки: [zwn2015@yandex.ru](mailto:zwn2015@yandex.ru)

**ЮДИН Евгений Борисович**, кандидат технических наук, старший научный сотрудник омского филиала Института математики им. С. Л. Соболева СО РАН. Адрес для переписки: [udinev@asoiu.com](mailto:udinev@asoiu.com)

**ЮДИНА Мария Николаевна**, аспирантка кафедры «Автоматизированные системы обработки информации и управления» ОмГТУ.

Адрес для переписки: [mg-and-all@mail.ru](mailto:mg-and-all@mail.ru)

Статья поступила в редакцию 25.09.2017 г.

© В. Н. Задорожный, Е. Б. Юдин, М. Н. Юдина

## ДОВЕРИТЕЛЬНЫЕ ИНТЕРВАЛЫ ДЛЯ КВАНТИЛЕЙ РАСПРЕДЕЛЕНИЯ

В работе рассматривается построение асимптотических доверительных интервалов для квантилей параметрических, непараметрических и семинепараметрических классов распределений с учетом различной априорной информации об исходном распределении. В качестве оценки квантиля неизвестной функции распределения используется решение эмпирического уравнения. Для решения уравнения используются рекуррентные методы стохастической аппроксимации. Показано, что асимптотические доверительные интервалы для квантилей распределения полностью определяются через оценки функции распределения и их дисперсии. В общем виде приведены асимптотические доверительные интервалы для квантилей. Выведены оценки доверительных интервалов для квантилей с учетом априорной информации об исходном распределении.

**Ключевые слова:** квантиль, доверительные интервалы, семинепараметрические интервалы, непараметрические интервалы.

**Введение.** В различных областях приложений методов математической статистики (теория надежности, статистический контроль качества, разработка новых изделий и материалов, автоматизированное управление технологическими процессами ...) возникают задачи построения точечных и интервальных оценок для квантилей функции распределения случайных величин. В теории надежности — гамма-процентный ресурс определяется как квантиль функции распределения наработки изделия до отказа, в теории статистического контроля качества — доля дефектных деталей определяется как квантиль соответствующего уровня оперативной характеристики [1, 2]. Несмотря на хорошо разработанную теорию построения доверительных интервалов [3–5], следует отметить отсутствие общего подхода к построению доверительных интервалов для квантилей [5–7]. Данная задача решена для некоторых классов параметрических распределений (обзор в [7, 8]), для непараметрических классов распределений построены непараметрические доверительные интервалы на основе порядковых статистик [5–7]. Основные методы построения доверительных интервалов базируются на использовании точечных оценок, значительное количество которых имеет асимптотически нормальное распределение. На этом факте основан общий подход построения асимптотических доверительных интервалов. В данной работе рассматривается построение асимптотических доверительных интервалов для квантилей параметрических, непараметрических и семинепараметрических классов распределений с учетом различной априорной информации об исходном распределении имеющих свою специфику и область применений.

**1. Постановка задачи.** Пусть  $X$  — случайная величина с функцией распределения  $F(x) \subset \Omega$ , где  $\Omega$  класс параметрических или непараметрических непрерывных распределений,  $f(x)$  — плотность распределения  $X$ . Обозначим через  $X_p$  единственный квантиль распределения уровня  $p$  ( $0 < p < 1$ ), т.е. уравнение  $F(X_p) = p$  имеет единственное решение. Требуется по выборке  $\tilde{X}_N = (x_1, \dots, x_N)$  — объема  $N$  независимых и одинаково распределенных (н.о.р.) случайных величин построить оценку  $\tilde{X}_{pN}$  и доверительный интервал  $(\tilde{X}_{pN}^-, \tilde{X}_{pN}^+)$  для квантиля  $X_p$  для различных классов  $\Omega$  непрерывных распределений.

**2. Оценки и асимптотические доверительные интервалы для квантилей.** На основе метода подстановки [3, 4, 9] в качестве состоятельной оценки  $\tilde{X}_{pN}$  квантиля  $X_p$  функции распределения  $F(x)$  возьмем решение эмпирического уравнения

$$\hat{F}_N(\tilde{X}_{pN}) = p \Leftrightarrow \tilde{X}_{pN} = \hat{F}_N^{-1}(p), \quad (1)$$

где  $\hat{F}_N(x)$  — состоятельная и несмещенная (асимптотически несмещенная) оценка  $F(x)$ .

В дальнейшем ограничимся достаточно общим классом асимптотически нормальных оценок  $\hat{F}_N(x)$ , т.е. будем предполагать, что случайная величина

$$\sqrt{N}[\hat{F}_N(x) - F(x)] \Leftrightarrow \Phi(0, \sigma^2)$$

имеет асимптотически нормальное распределение  $\Phi(0, \sigma^2)$  с нулевым средним и дисперсией  $\sigma^2$

$$\sigma^2 = \sigma^2(\hat{F}_N(x)) = \sigma^2(F(x), \bar{\theta}). \quad (2)$$

В ряде параметрических задач  $F(x, \bar{\theta}) \subset \Omega_p$  оценка  $\tilde{X}_{pN}$  может быть выражена явно, например, для

экспоненциального распределения [3, 4]. В общем случае для нахождения оценки  $\hat{X}_{pN}$  используются рекуррентные методы стохастической аппроксимации вида

$$\hat{X}_{pN}[k+1] = \hat{X}_{pN}[k] - \gamma[k] \left[ \hat{F}_N(\hat{X}_{pN}[k]) - p \right], \quad k=0, 1, \dots, \quad (3)$$

где  $\gamma[k]$  должны удовлетворять условиям  $\gamma[k] > 0$ ;

$$\gamma[k] > 0; \sum_{k=0}^{\infty} \gamma[k] = \infty; \sum_{k=0}^{\infty} \gamma^2[k] < \infty.$$

Рассмотрим свойства оценки квантиля  $\hat{X}_{pN}$ . Для этого представим (1) в виде [3, 9]

$$[\hat{X}_{pN} - X_p] = \frac{1}{f(X_p)} [\hat{F}_N(\hat{X}_{pN}) - F(X_p)]. \quad (4)$$

**Теорема 1.** Если  $\sqrt{N}[\hat{F}_N(x) - F(x)] \Leftrightarrow \Phi(0, \sigma^2)$  и  $\sigma^2$  из (2) непрерывная функция  $F(x)$ , то  $\sqrt{N}[\hat{X}_{pN} - X_p] \Leftrightarrow \Phi(0, D(\hat{X}_{pN}))$ , где

$$D(\hat{X}_{pN}) = \frac{\sigma^2}{f^2(X_p)}. \quad (5)$$

Доказательство прямо следует из представления (4) и теорем непрерывности [3].

Пусть

$$\hat{D}_N(\hat{X}_{pN}) = [f_N(X_p)]^{-2} \hat{\sigma}_N^2 \quad (6)$$

состоятельная и непрерывная оценка дисперсии оценки квантиля  $D(\hat{X}_{pN})$ .

**Теорема 2.** Если  $D(\hat{X}_{pN})$  непрерывная функция  $F(x)$  и  $\hat{D}_N(\hat{X}_{pN})$  состоятельная и непрерывная оценка  $D(\hat{X}_{pN})$ , то

$$\sqrt{N} \frac{[\hat{X}_{pN} - X_p]}{\sqrt{\hat{D}_N(\hat{X}_{pN})}} \Leftrightarrow \Phi(0, 1). \quad (7)$$

Доказательство следует из второй теоремы непрерывности [3].

Теоремы 1, 2, являющиеся частным случаем общего подхода к построению асимптотических доверительных интервалов [3–7], дают основу для построения асимптотических интервалов для квантилей. Действительно, из (7) следует, что можно записать

$$\Pr \left\{ \sqrt{N} \frac{[\hat{X}_{pN} - X_p]}{\sqrt{\hat{D}_N(\hat{X}_{pN})}} < \beta \right\} = 1 - \varepsilon.$$

Это соотношение можно переписать в виде

$$\Pr \left\{ \sqrt{N} [\hat{X}_{pN} - \beta \cdot \sqrt{\hat{D}_N(\hat{X}_{pN})}] < X_p < \sqrt{N} [\hat{X}_{pN} + \beta \cdot \sqrt{\hat{D}_N(\hat{X}_{pN})}] \right\} = 1 - \varepsilon,$$

где  $\beta = \beta(\varepsilon)$  — квантиль уровня  $(0,5 \cdot \varepsilon)$  стандартного нормального распределения [3, 4]. В результате получаем асимптотические доверительные интервалы для квантилей в виде

$$\left[ \hat{X}_{pN} - \frac{1}{\sqrt{N}} \beta \cdot \sqrt{\hat{D}_N(\hat{X}_{pN})}, \right.$$

$$\left. \hat{X}_{pN} + \frac{1}{\sqrt{N}} \beta \cdot \sqrt{\hat{D}_N(\hat{X}_{pN})} \right]. \quad (8)$$

Из (8) с учетом (6) получаем, что длина  $L$  доверительного интервала будет

$$L = \frac{2}{\sqrt{N}} \beta \cdot \sqrt{\hat{D}_N(\hat{X}_{pN})} = \frac{2}{\sqrt{N}} \beta \cdot [f_N(X_p)]^{-1} \hat{\sigma}_N. \quad (9)$$

Пусть имеются две оценки  $\hat{X}_{pN}(1)$  и  $\hat{X}_{pN}(2)$ , тогда эффективность оценки доверительного интервала вида (9) на основе  $\hat{X}_{pN}(1)$  по отношению к оценке доверительного интервала вида (9) на основе  $\hat{X}_{pN}(2)$ , с учетом, что  $\hat{\sigma}_N^2(\hat{F}_{Nj}(x))$  состоятельная оценка  $\sigma^2(\hat{F}_{Nj}(x))$  можно определить в виде отношения

$$\text{eff} [\hat{X}_{pN}(1), \hat{X}_{pN}(2)] = \frac{[L(\hat{X}_{pN}(2))]^2}{[L(\hat{X}_{pN}(1))]^2} = \frac{\sigma^2(\hat{F}_{N2}(x))}{\sigma^2(\hat{F}_{N1}(x))}, \quad (10)$$

где  $\sigma^2(\hat{F}_{Nj}(x))$  — дисперсия оценки  $j$ -ой функции распределения.

Как следует из (8), (9) асимптотические доверительные интервалы для квантилей распределения полностью определяются через оценки функции распределения и их дисперсии.

**3. Оценивание функций распределения.** Рассмотрим задачу оценивания функции распределения  $X$  при разных уровнях априорной информации о случайной величине.

**3.1. Параметрические оценки.** Пусть  $F(x, \bar{\theta}) \subset \Omega_p$  — параметрическая функция распределения, определенная с точностью до конечного числа неизвестных параметров  $\bar{\theta} = (\theta_1, \dots, \theta_k)^T$ ,  $\Omega_p$  класс параметрических функций распределения и  $\bar{\theta}_N = (\theta_{1N}, \dots, \theta_{kN})^T$  — состоятельная оценка параметра  $\bar{\theta} = (\theta_1, \dots, \theta_k)^T$ . Для построения состоятельной оценки  $F(x, \bar{\theta})$  воспользуемся методом подстановки [3, 4, 9]. В качестве оценки  $F(x, \bar{\theta})$  возьмем  $\hat{F}_N(x, \bar{\theta}) = F(x, \bar{\theta}_N)$ . **Теорема 3.** Пусть  $F(x, \bar{\theta})$  непрерывная функция по  $\theta$  и имеет непрерывные и ограниченные производные по  $x$  до третьего порядка,  $\sqrt{N}(\bar{\theta}_N - \bar{\theta}) \Leftrightarrow \Phi(0, \mathbf{B})$ ,  $\mathbf{B}$  — ковариационная матрица  $\bar{\theta}_N = (\theta_{1N}, \dots, \theta_{kN})^T$ , тогда

$$\sqrt{N} [\hat{F}_N(x, \bar{\theta}_N) - F(x, \bar{\theta})] \Leftrightarrow \Phi(0, D\hat{F}_N(x, \bar{\theta}_N)), \quad (11)$$

где

$$D\hat{F}_N(x, \bar{\theta}_N) = C^T \mathbf{B} C, \quad (12)$$

$$C^T = \left( \frac{\partial F(x, \theta)}{\partial \theta_1}, \dots, \frac{\partial F(x, \theta)}{\partial \theta_k} \right). \quad (13)$$

Доказательство.

Из условий теоремы следует, что имеет место представление

$$[\hat{F}_N(x, \bar{\theta}_N) - F(x, \bar{\theta})] = C^T (\bar{\theta}_N - \bar{\theta})$$

и доказательство следует из третьей теоремы непрерывности [3].

В [7] на основе (11) получены оценки и доверительные интервалы для квантилей ряда параметрических распределений.

**3.2. Непараметрические оценки.** Пусть  $F(x) \subset \Omega_n$  — непараметрическая функция распределения, вид которой неизвестен, кроме некоторых общих предположений типа непрерывности, симметричности и т.д., и  $\Omega_n$  класс непараметрических функций распределения. Классической непараметрической

оценкой  $F(x) \subset \Omega_n$  является эмпирическая функция распределения (э.ф.р.)  $F_N(x)$

$$F_N(x) = \frac{1}{N} \sum_{i=1}^N C(x - x_i), \quad (14)$$

где

$$C(x) = \begin{cases} 1, & x \geq 0, \\ 0, & x < 0. \end{cases} \quad (15)$$

Эмпирическая функция распределения  $F_N(x)$  является несмещенной и эффективной оценкой  $F(x)$  на классе  $\Omega_n$  и

$$\sqrt{N}[F_N(x) - F(x)] \Leftrightarrow \Phi(0, \sigma^2), \quad (16)$$

где

$$\sigma^2 = F(x)[1 - F(x)]. \quad (17)$$

Решение уравнения (1) вида  $\hat{X}_{pN} = F_N^{-1}(p)$  приводит к непараметрическим оценкам квантилей и доверительным интервалам на основе порядковых статистик [5, 6], но дискретный характер э.ф.р. доставляет определенные трудности при решении уравнения (1). Дискретный характер э.ф.р. определяет функция  $C(t)$ , определенная в (15), поэтому можно сгладить функцию  $C(t)$  [3, 6, 9]. Например, заменить «ступеньки лестницы» э.ф.р. прямыми. Действительно, пусть  $X$  определена на  $[a, b]$ , т.е.  $F(a) = 0$  и  $F(b) = 1$ . Обозначим через

$$a = x_{(0)} < x_{(1)} < x_{(2)} < \dots < x_{(N)} < x_{(N+1)} = b$$

— расширенный вариационный ряд.

Заменяя «ступеньки лестницы» э.ф.р. прямыми, получим следующую оценку

$$\tilde{F}_N(y) = \begin{cases} \frac{1}{2N} \left( \frac{y - x_{(0)}}{x_{(1)} - x_{(0)}} \right), & \text{для } y < x_{(1)} \\ \frac{1}{2N} \left( (2i - 1) + \frac{y - x_{(i)}}{x_{(i+1)} - x_{(i)}} \right), & \text{для } y \in [x_{(i)}, x_{(i+1)}], \\ \left( 1 + \frac{y - x_{(N+1)}}{2N(x_{(N+1)} - x_{(N)})} \right), & \text{для } y > x_{(N)} \end{cases} \quad (18)$$

Оценка (18) уже будет оценкой в классе непрерывных оценок, но не очень гладкой. Чтобы повысить степень гладкости оценки, используется приближение в виде монотонно не убывающей и дифференцируемой функции  $G(u)$ , полинома или сплайна. В результате получим следующую оценку

$$\tilde{F}_N(y) = \begin{cases} \frac{1}{2N} \left( \frac{G(y) - G(x_{(0)})}{G(x_{(1)}) - G(x_{(0)})} \right), & \text{для } y < x_{(1)} \\ \frac{1}{2N} \left( (2i - 1) + \frac{G(y) - G(x_{(i)})}{G(x_{(i+1)}) - G(x_{(i)})} \right), & \text{для } y \in [x_{(i)}, x_{(i+1)}], \\ \left( 1 + \frac{G(y) - G(x_{(N+1)})}{2N[G(x_{(N+1)}) - G(x_{(N)})]} \right), & \text{для } y > x_{(N)}. \end{cases} \quad (19)$$

К сожалению, такой подход оказался не очень конструктивным, так как при его обобщении на многомерный случай возникают определенные трудности математического характера. Функция  $C(t)$  относится к классу обобщенных функций, которые

в теории представляются как предел последовательностей гладких аналитических дифференцируемых функций. Данное представление  $C(t)$  легко обобщается на многомерный случай и позволяет рассматривать различные эмпирические подходы к построению оценок э.ф.р., типа рассмотренных выше с единых позиций. Обозначим через  $\tilde{C}(t)$  аналитическое приближение к  $C(t)$ .

В частности,  $C(t)$  можно представить в виде

$$\tilde{C}(t, h) = G\left(\frac{t}{h}\right), \quad (20)$$

где  $h$  — параметр размытости и  $G(t)$  — непрерывная и симметричная функция распределения.

Свяжем параметр  $h$  с объемом выборки так, чтобы выполнялось условие  $h_N \rightarrow 0$ , при  $N \rightarrow \infty$ . Можно показать [9], что если  $h_N \rightarrow 0$ , то  $\tilde{C}(t, h_N) \rightarrow C(t)$  в обычном смысле. Тогда сглаженную эмпирическую функцию распределения можно определить в виде [3], [9]

$$\tilde{F}_N(x) = \frac{1}{N} \sum_{i=1}^N G\left(\frac{x - x_i}{h_N}\right). \quad (21)$$

Можно показать [9], что при  $h_N \rightarrow 0$ ,  $N \rightarrow \infty$   $\tilde{F}_N(x)$  является непрерывной, асимптотически несмещенной, состоятельной оценкой  $F(x)$  и  $\sqrt{N}(\tilde{F}_N(x) - F(x)) \Rightarrow \Phi(0, F(x)[1 - F(x)])$  т.е.  $\tilde{F}_N(x)$  является асимптотически эффективной непараметрической оценкой  $F(x)$  на непараметрическом уровне априорной информации.

В [7, 8] на основе (21) получены непараметрические оценки и доверительные интервалы для квантилей.

**3.3. Семинепараметрические оценки.** Располагая некоторой дополнительной априорной информацией о  $F(x) \subset \Omega_{np}$ , где  $\Omega_{np}$  — класс семинепараметрических распределений, можно построить оценки функции распределения (ф.р.) с учетом этой априорной информации, свойства которых лучше, чем у э.ф.р. [9, 10].

Пусть априори известно, что  $F(x)$  удовлетворяет условиям

$$S_i = \int \psi_i(x) dF(x) = 0, \quad i = 1, \dots, r, \quad (22)$$

где функции  $\psi_1, \dots, \psi_r$  — известны. Обозначим через

$$S_{ij} = \int \psi_i(x) \psi_j(x) dF(x) < \infty, \quad i, j = 1, \dots, r$$

и образуемая ими матрица  $\Lambda = \|S_{ij}\|$  не вырождена.

Определим  $\Omega_{np}$  — семинепараметрический класс непараметрических распределений с учетом априорной информации вида (22) и положим  $F(x) \subset \Omega_{np}$ . Знание априорной информации в виде (22) позволяет учесть широкий спектр информации о  $F(x)$ , как количественного, так и качественного характера. В [10, 11] был предложен метод построения модифицированных оценок  $F(x) \subset \Omega_{np}$  вида

$$\bar{F}_N(x) = F_N(x) - \left\| \int \psi_i(y) dF_N(y) \right\|^T \Lambda_N^{-1} \left\| \int C(x - y) \psi_j(y) dF_N(y) \right\|, \quad (23)$$

При некоторых условиях регулярности можно доказать [11], что  $\bar{F}_N(x)$  является асимптотически несмещенной, состоятельной и эффективной оценкой для  $F(x) \subset \Omega_{np}$ , причем  $\sqrt{N}[\bar{F}_N(x) - F(x)] \Leftrightarrow \Phi(0, \sigma^2)$  с дисперсией

$$\sigma^2 = F(x) - F^2(x) -$$



$$- \left\| \int C(x-y)\psi_i(y)dF(y) \right\|^T \Lambda^{-1} \left\| \int C(x-y)\psi_j(y)dF(y) \right\|. \quad (24)$$

Анализ (24) показывает, что величина

$$\left\| \int C(x-y)\psi_i(y)dF(y) \right\|^T \Lambda^{-1} \left\| \int C(x-y)\psi_j(y)dF(y) \right\| \geq 0,$$

и является неубывающей функцией по  $r$ . Следовательно, введение дополнительной априорной информации может лишь улучшить свойства оценки  $\bar{F}_N(x)$ ,  $D\bar{F}_N(x) \leq DF_N(x)$  по сравнению с э.ф.р.

Рассмотрим ряд примеров.

**Пример 1.** Пусть известно, что среднее значение равно нулю, т. е. имеем  $r = 1$ ,  $\psi(x) = x$ . Тогда модифицированная оценка ф.р. (23) примет вид

$$\bar{F}_{1N}(x) = F_N(x) - \frac{\int x dF_N(x)}{\int x^2 dF_N(x)} \int C(x-y)y dF_N(y), \quad (25)$$

с дисперсией

$$\sigma_1^2 = [F(x) - F^2(x)] - \frac{1}{\sigma^2} \left[ \int C(x-y)y dF(y) \right]^2 < [F(x) - F^2(x)]. \quad (26)$$

**Пример 2.** Пусть априори известно, что  $F(x)$  — симметричная функция относительно точки  $\theta$ , т. е.

$$\psi(y, x, \theta) = \frac{1}{2} [C(x-y) + 1 - C(2\theta - x - y)].$$

Тогда модифицированная оценка ф.р. (23) примет вид

$$\bar{F}_{2N}(x) = \frac{1}{2} [F_N(x) + 1 - F_N(2\theta - x)]. \quad (27)$$

Дисперсия оценки будет

$$\sigma_2^2 = \frac{1}{2} F(x)[1 - 2F(x)], \quad \text{если } x < 0; \quad (28)$$

$$\sigma_2^2 = \frac{1}{2} [1 - F(x)][2F(x) - 1], \quad \text{если } x \geq 0.$$

По аналогии с подразделом (23) можно ввести smoothed модифицированную эмпирическую функцию распределения [9]

$$\tilde{F}_N(x) = \bar{F}_N(x) - \left\| \int \psi_i(y)dF_N(y) \right\|^T \Lambda_N^{-1} \left\| \int G\left(\frac{x-y}{h_N}\right)\psi_j(y)dF_N(y) \right\|. \quad (29)$$

#### 4. Оценка доверительного интервала квантиля.

Как следует из (8), для нахождения доверительного интервала квантиля необходимо определить следующие величины — оценку квантиля  $\hat{X}_{pN}$  из уравнения (1) и оценку дисперсии оценки квантиля  $\hat{D}_N(\hat{X}_{pN})$  на основе априорной информации о распределении  $F(x) \subset \{\Omega_p, \Omega_{pn}, \Omega_n\}$ . Параметры  $X_p$  и  $D(\hat{X}_{pN})$  являются функционалами от функции распределения и их оценивают в соответствии с методом подстановки [9] на основе введенных в разделе 3 оценок ф.р.

**5. Заключение.** В работе предложен общий подход к построению доверительных интервалов для квантилей распределения исходя из априорной информации, имеющейся о распределении. Рассмотрены классы параметрических, непараметрических и семинепараметрических распределений. Отметим, что для каждого уровня априорной информации строятся асимптотически эффективные решения. Данный подход позволяет расширить область принимаемых решений и строить доверительные интервалы для классов семинепараметрических и непараметрических распределений, учитывающих различного рода выбросы.

#### Библиографический список

1. Павлов И. В. Статистические методы оценки надежности сложных систем по результатам испытаний. М.: Радио и связь, 1982. 168 с.
2. Надежность и эффективность в технике. В 10 т. / Редак. совет: В. С. Авдуевский, В. И. Кузнецов, Н. Д. Кузнецов [и др.]. М.: Машиностроение, 1986. Т. 2. 280 с.
3. Боровков А. А. Математическая статистика. М.: Наука, 1984. 472 с.
4. Шуленин В. П. Математическая статистика. В 3 ч. Ч. 1. Параметрическая статистика. Томск: Изд-во НТЛ, 2012. 539 с. ISBN 978-5-89503-492-7.
5. Уилкс С. Математическая статистика. М.: Наука, 1967. 632 с.
6. Шуленин В. П. Математическая статистика В 3 ч. Ч. 2. Непараметрическая статистика. Томск: Изд-во НТЛ, 2012. 387 с. ISBN 978-5-89503-502-3.
7. Симахин В. А., Терещенко Е. Р. Доверительные интервалы для квантилей и функций распределения. Курган, 1988. 55 с. Деп. в ВИНТИИ, № 7145-888.
8. Tereshchenko E. P., Teskin O. I., Simakhin V. A. Bootstrap models for interval estimation of longevity characteristics of sequential systems from small samples // Journal of Mathematical Sciences. 1996. Vol. 81, Issue 4. P. 2811–2817.
9. Симахин В. А. Робастные непараметрические оценки. Saarbrücken: LAP Lambert Academic Publishing, 2011. 300 с. ISBN 978-3-8465-5771-6.
10. Тюрин Ю. Н. Об оценивании функции распределения // Теория вероятностей и ее применения. 1970. Т. 15, № 3. С. 549–550.
11. Дмитриев Ю. Г. О свойствах оценок функций распределения и функционалов при дополнительной априорной информации // Математическая статистика и ее приложения. 1976. № 4. С. 63–76.

**БАТРАКОВ Петр Андреевич**, кандидат технических наук, доцент кафедры «Теплоэнергетика» Омского государственного технического университета.

**МАЕР Алексей Владимирович**, кандидат технических наук, доцент кафедры «Программное обеспечение автоматизированных систем (ПОАС)» Курганского государственного университета (КГУ).

**СИМАХИН Валерий Ананьевич**, кандидат физико-математических наук, профессор кафедры «Программное обеспечение автоматизированных систем (ПОАС)» КГУ.

Адрес для переписки: alex\_povt@mail.ru

Статья поступила в редакцию 29.09.2017 г.

© П. А. Батраков, А. В. Маер, В. А. Симахин

Омский государственный  
технический университет,  
г. ОмскЕвразийский национальный университет  
им. Л. Н. Гумилева,  
г. Астана, Республика КазахстанКазахстанский филиал  
Московского государственного  
университета  
им. М. В. Ломоносова,  
г. Астана, Республика Казахстан

## МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ДИНАМИКИ РОСТОВЫХ ПРОЦЕССОВ КАРТОФЕЛЯ В УСЛОВИЯХ IN VITRO ПРИ ИСПОЛЬЗОВАНИИ РАЗЛИЧНЫХ ВАРИАНТОВ ПИТАТЕЛЬНЫХ СРЕД НА ОСНОВЕ ДИНАМИЧЕСКИХ МОДЕЛЕЙ ОБРАБОТКИ РЯДОВ

В настоящее время использование математических методов моделирования для оптимизации биологических процессов является актуальным и перспективным направлением. Метод математического моделирования нашел свое применение в моделировании ростовых процессов растений при микроклональном размножении в зависимости от различных факторов культивирования. Целью работы было изучение влияния концентрации фитогормонов и витаминов на процессы морфогенеза растений-регенерантов сибирских сортов картофеля ранних «Ермак», «Алена» и среднераннего «Хозяюшка» в условиях *in vitro*. В работе представлено исследование по выбору наиболее эффективного компонентного состава питательной среды для микроклонального размножения картофеля сибирских сортов картофеля «Ермак», «Алена», «Хозяюшка». В результате исследований было установлено, что лучшим вариантом среды для культивирования микропобегов изучаемых сортов картофеля является питательная среда с минеральным составом по прописи Мурасиге—Скуга с внесением кинетина 0,5 мг/л, ИУК 0,2 мг/л, тиамин 1,5 мг/л, пиридоксин 1,0 мг/л, аскорбиновой кислоты 3,0 мг/л, феруловой кислоты 0,005 мг/л, сахарозы 30000 мг/л.

**Ключевые слова:** микроклональное размножение, аппроксимация, сингулярный спектральный анализ, питательные среды, сибирские сорта картофеля.

Основным этапом в микроклональном размножении растений является подбор состава питательной среды для ускорения роста и развития растений-регенерантов, особенно это касается высоты растений в культуре *in vitro*, которая напрямую влияет на количество получаемых микропобегов и увеличивает коэффициент размножения.

**Цель работы** — математическое моделирование влияния концентраций фитогормонов и витаминов на процессы морфогенеза растений-регенерантов картофеля сортов «Ермак», «Алена», «Хозяюшка» в условиях *in vitro*.

**Материалы и методы.** В данной работе представлено исследование по выбору наиболее эффектив-

ного состава питательной среды для микроклонального размножения картофеля сибирских сортов картофеля «Ермак», «Алена», «Хозяюшка» с использованием математического моделирования динамики ростовых процессов. Изучались показатели ростовых процессов: высота растения-регенеранта и корневая система в зависимости от времени культивирования (7, 14, 21 сутки). Измерения проводили в 3-х повторностях.

Использованная в составе сред феруловая кислота является важным биологическим компонентом растений, обладает сильным действием на ростовые процессы, выступая в качестве активатора ИУК-оксидазы, обладает ингибирующим действием против

вирусов растений [1]. Улановым А. В. было установлено, что развитие растительных тканей *in vitro* связано с изменениями содержания фенольных соединений клеточной стенки. Содержание феруловой кислоты в клеточных стенках регенерирующих культур значительно выше по сравнению с не регенерирующими культурами, что может быть связано с формированием определенной морфологической структуры тканей, способных к регенерации и стимулированию ростовых процессов [2].

Использованию математических методов моделирования биологических процессов на сегодняшний день посвящено достаточно большое количество работ, наиболее близкими из которых к рассматриваемым в статье вопросам нашли отражение в работах многих отечественных и зарубежных ученых [3–5]. В работах Лебедевой Н. В. [6] и Ковалева А. И. [7] был использован регрессионный анализ для определения оптимальных параметров выращивания растений картофеля *in vitro*. В работе при разработке математической модели использовались методы сингулярного спектрального анализа (SSA) и аппроксимации (LA), которые дают возможность прогнозирования данных биологических процессов [8–11].

**Постановка и решение задачи.** При математическом моделировании определялось оптимальное соотношение гормональных факторов и витаминов при культивировании растений-регенерантов сибирских сортов картофеля. При анализе различных рядов данных главной задачей являлась реконструкция динамической системы. В соответствии с теорией Такенса–Мане [12] приемлемое описание фазового пространства динамической системы можно получить, если взять вместо реальных переменных системы, которые могут быть неизвестны,  $k$ -мерные векторы задержек, составленные из значений ряда в последовательные моменты времени.

При выполнении условия  $k > 2d_e + 1$ , где  $d_e$  — размерность вложений, возможно реконструировать фазовое пространство системы. При условии стационарности ряда на базе этой реконструкции строится прогноз его дальнейшей динамики. В силу того, что ряды достаточно малы, наиболее оптимальным методом для определения величины —  $k$  является функциональный метод. Согласно использованию метода Грассбергера–Прокачия [13], необходимо построить через равные промежутки времени  $\tau$  последовательность  $y_1, y_2, \dots, y_n$ , наблюдаемой величины  $x(t)$  следующим образом:

$$x(t) = y_1, x(t + \tau) = y_2, \dots, x(t + (n-1)\tau) = y_n. \quad (1)$$

Далее выбрался из имеющейся последовательности, начиная с некоторого номера  $n$ , а также произвольной подпоследовательности размера  $k$ , и рассматривались всевозможные векторы значений наблюдаемой величины:

$$w^n = (y_{n-k+1}, y_{n-k+2}, \dots, y_{n-1}, y_n). \quad (2)$$

Допустим, что динамика исходной рассматриваемой системы описывается одним дифференциальным уравнением первого порядка. Тогда эта последовательность будет определяться единственным начальным значением  $y_1$ , а все остальные  $y_n, n > 1$ , будут функционально зависеть от него:

$$y_{n+1} = f(w^n). \quad (3)$$

Наиболее интересным для решения поставленной задачи будет такое  $k$ , при котором это расстояние

будет стремиться к нулю, где размерность вложения  $d_e$  — это минимальное значение  $k$ . Значение  $k$  — размер вектора задержек,  $d_e$  — размерность вложений, возможно, реконструировать фазовое пространство системы. Значение  $\tau$  — равные промежутки времени,  $r(n, n_0) = |f(w^n) - f(w^0)|$  — расстояние между векторами задержек.

На основании метода сингулярного спектрального анализа идет преобразование полученного скалярного ряда, содержащего  $n$  значений наблюдаемой величины, в матрицу задержек, в которой  $y_k, y_n$  — это значения наблюдаемой величины, которые составляют элементы вектора последовательности  $w^n$ :

$$(y_1, y_2, \dots, y_n) \rightarrow Y_k \cdot (n - K + 1) = \begin{pmatrix} y_k & y_{k+1} & \dots & y_n \\ \vdots & \vdots & \vdots & \vdots \\ y_2 & y_3 & \dots & y_{n-k+2} \\ y_1 & y_2 & \dots & y_{n-k+1} \end{pmatrix}. \quad (4)$$

Размерность матрицы определяется количеством задержек  $k$ . В ходе построения матрицы задержек выбирается вид локального представления, т.е. вид функции, связывающей следующее значение ряда с предыдущими:

$$y_{t+1} = f(y_t, a), \quad (5)$$

где  $a$  — вектор параметров представления, а  $y_t$  — это значения наблюдаемой величины.

В данном случае наиболее удобным вариантом является линейная аппроксимация первого порядка:

$$y_{t+1} = a_0 + ya. \quad (6)$$

Далее выделяется локальная подобласть фазового пространства, в которой будет строиться прогноз данных. При решении данной задачи прогноз будет строиться на основе критерия вида:

$$|y_{n-k+1} - y_s| < \varepsilon, \quad (7)$$

где  $\varepsilon$  — является критерием выбора подобласти, это достаточно малое число, приблизительно равное 0,001,  $y_s$  — это некоторое значение наблюдаемой величины.

Параметры модели, а именно, вектор  $a$ , будет оцениваться методом наименьших квадратов (МНК). Оценка по методу наименьших квадратов для вектора  $a$ , обозначенного как  $\hat{a}$ , находится из следующих условий:

$$\hat{a}: \sum_s (y_{s+1} - f(y_s, \hat{a}))^2 \rightarrow \min. \quad (8)$$

Оценив значения аппроксимации, строится прогноз следующего значения ряда  $y_{t+1} = f(y_t, \hat{a})$ , которое является результатом прогнозирования биотехнологического процесса.

**Программная реализация разработанных алгоритмов.** Математическую обработку экспериментальных данных, осуществляли в программе Scilab 5.5.2, разработанной на языке программирования C#, который выбирался с требованием максимальной переносимости комплекта программ под различные операционные системы [14].

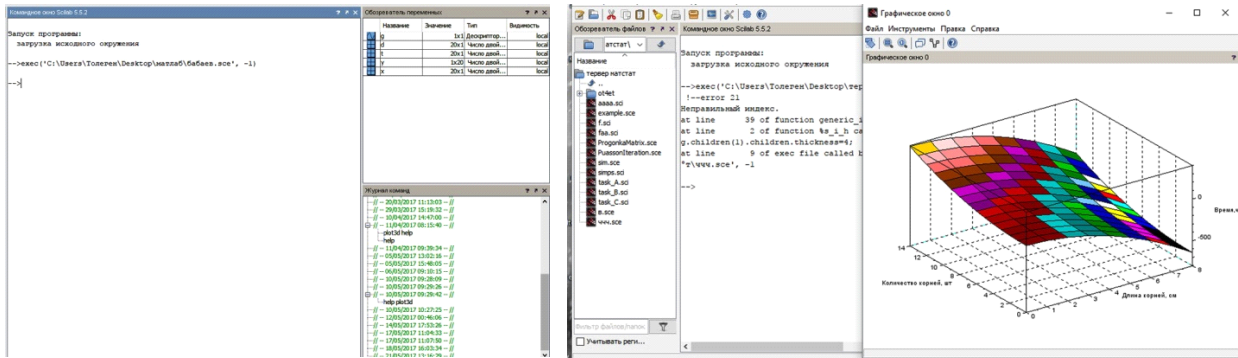


Рис. 1. Интерфейс программы Scilab 5.5.2, в которой выполнялось построение графиков

Для создания графического интерфейса был использован Microsoft Visual Studio. Интерфейс программы Scilab 5.5.2 и результаты вычислений представлены на рис. 1.

Ниже представлен фрагмент программного кода, реализующий формулу вычисления:

$$\hat{a} : \sum_{W_s} (y_{s+1} - f(y_s, \hat{a}))^2 \rightarrow \min .$$

```

Clear function [y_out, m, sig] = ssa_normalize(y)
    m = mean(y)
    sig = sqrt(variance(y)) // stdev(y)
    y_out = (y - m) / sig
endfunction
function [y_out] = ssa(y, L, I)
    [LAMBDA, U, V] = ssa_decompose(y, L)
    y_out = ssa_reconstruct(LAMBDA, U, V, I)
endfunction function [LAMBDA, U, V, X] =
ssa_decompose(y, L)
    // Ensure y is a column vector
    y = matrix(y, - 1, 1)
    // Stage 1: Decomposition
    // Step 1: Embedding
    N = length(y)
    if L > N / 2 then L = N - L
    end K = N - L + 1

```

```

// Time- delayed embedding of y, the tra-
jectory matrix
X = zeros(L, K)
for i = 1 : K
    X(:, i) = y(i : i + L - 1)
end // Step 2: Singular value decomposition
// Direct approach
//[U, LAMBDA] = svd(X)
//LAMBDA = diag(LAMBDA)
// Covariance matrix
C = X * X' / K
// For stationary series
//C = toeplitz(corr(X, L))
[U, LAMBDA] = svd(C)
// The eigenvalues of C are the squared eigen-
values of X
LAMBDA = sqrt(diag(LAMBDA))
// Principal components
V = X' * U
for i = 1 : L
    V(:, i) = V(:, i) / LAMBDA(i)
end endfunction
function [y] = ssa_reconstruct(LAMBDA, U, V, I)
// Ensure I is a row vector
I = matrix(I, 1, - 1)

```

Таблица 1

Различные варианты компонентного состава питательной среды Мурасиге – Скута (МС) для микроклонального размножения *in vitro*

Компоненты	Варианты питательной среды Мурасиге – Скута (МС) (мг/л)			
	МС (контроль)	МС-1	МС-2	МС-3
Минеральные компоненты	По МС	По МС	По МС	По МС
Тиамин	0,1	0,5	1,0	1,5
Пиридоксин	0,5	0,5	1,0	1,0
Аскорбиновая кислота	–	1,0	2,0	3,0
Мезоинозит	100	–	–	–
Глицин	2,0	–	–	–
Феруловая кислота	–	0,001	0,002	0,005
Кинетин	–	0,25	0,5	0,5
ИУК	–	0,05	0,1	0,2
Сахароза	30 000	20 000	25 000	30 000
Агар	7,0			
pH	5,6–5,8			

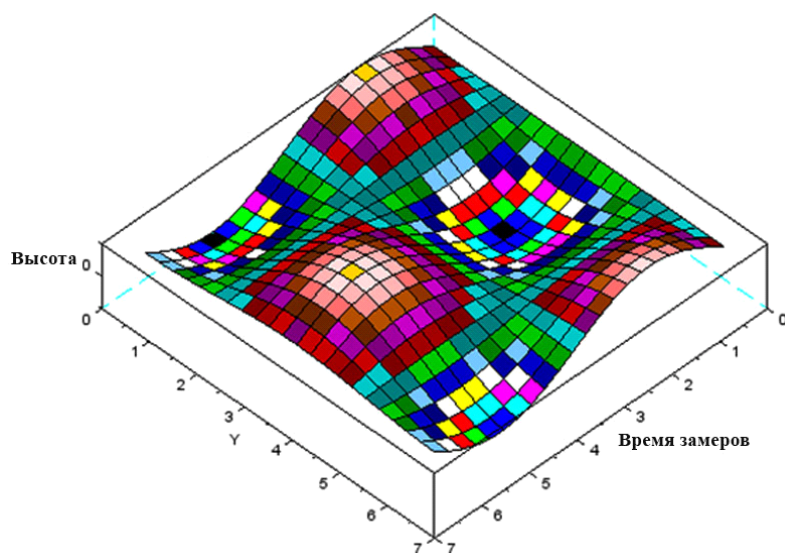


Рис. 2. Прогнозирование динамики ростовых процессов высоты растений-регенерантов сорта картофеля «Хозяюшка» *in vitro*

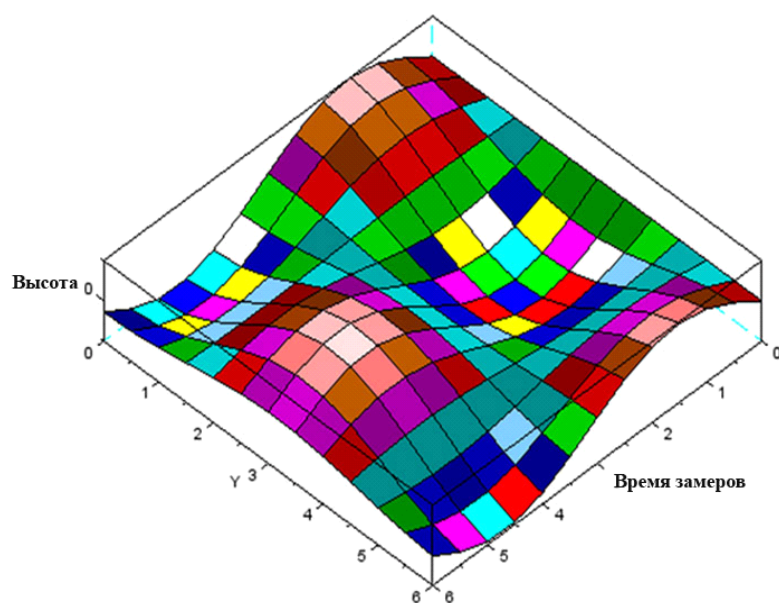


Рис. 3. Прогнозирование динамики ростовых процессов высоты растений-регенерантов сорта картофеля «Ермак» *in vitro*

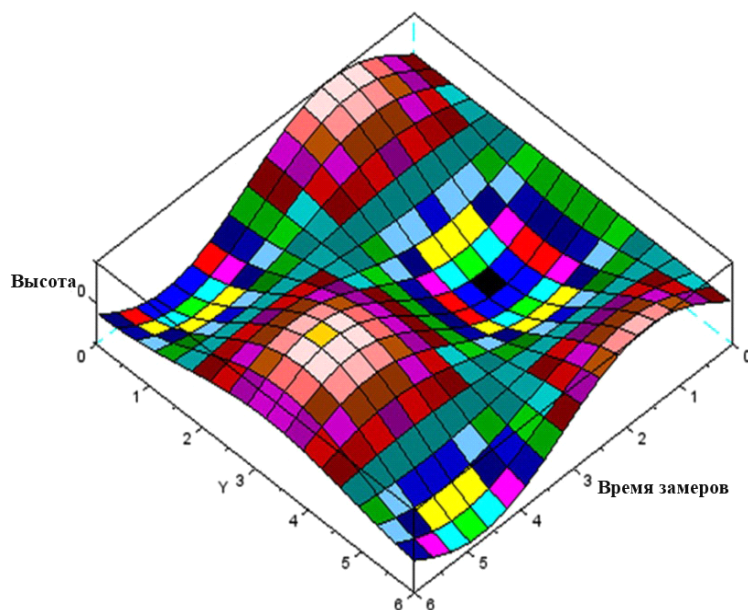


Рис. 4. Прогнозирование динамики ростовых процессов высоты растений-регенерантов сорта картофеля «Алена» *in vitro*

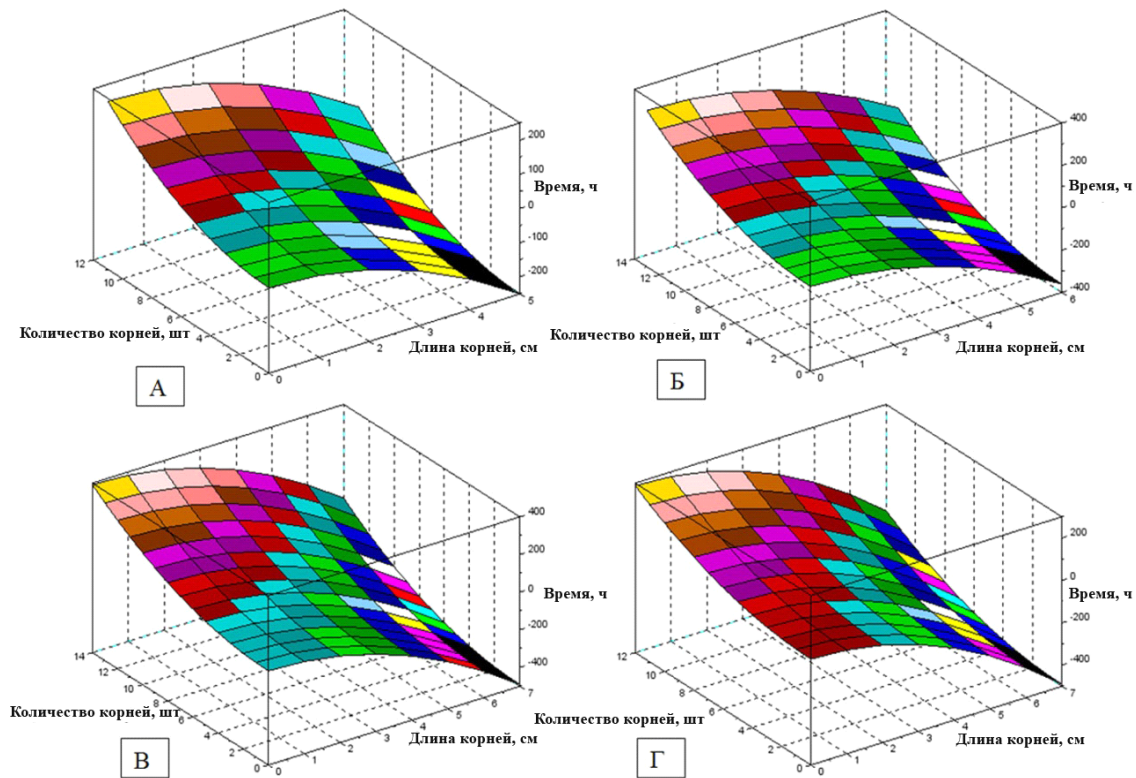


Рис. 5. Графики динамики ростовых процессов сорта картофеля «Хозяюшка» в условиях *in vitro*: А — среда МС (контроль); Б — среда МС-1; В — среда МС-2; Г — МС-3

$$[K, L] = \text{size}(V)$$

$$N = K + L - 1$$

При анализе результатов прогнозирования, культивирования растений ранних сортов картофеля «Ермак», «Алена» и среднераннего сорта «Хозяюшка» были построены графики зависимости высоты картофеля *in vitro*. На графиках время замеров ( $t$ ) измеряется в неделях культивирования, а высота растений картофеля ( $h$ ) — см.

Для микроклонального размножения картофеля использовали четыре варианта питательных сред по минеральной основе Мурасиге – Скуга с разным содержанием витаминов и фитогормонов (табл. 1).

**Результаты исследований.** При проведении микроклонального размножения раннеспелого сорта картофеля «Хозяюшка» на 7-е сутки среднее значение высоты на питательной среде (МС контроль) — 4,7; (МС-1) — 5,0; (МС-2) — 5,4; (МС-3) — 5,6; среднее значение высоты растений на 14-е сутки на среде (МС контроль) — 7,7; (МС-1) — 7,9; (МС-2) — 8,0; (МС-3) — 8,2; среднее значение высоты растений на 21-е сутки на среде (МС контроль) — 11,8; (МС-1) — 12,0; (МС-2) — 12,2; (МС-3) — 12,5.

При микроклональном размножении раннего сорта картофеля «Ермак» на 7-е сутки среднее значение высоты на среде (МС контроль) — 4,9; (МС-1) — 5,2; (МС-2) — 5,6; (МС-3) — 6,0; на 14-е сутки среднее значение высоты на среде (МС контроль) — 7,8; (МС-1) — 8,0; (МС-2) — 8,2; (МС-3) — 8,4; среднее значение высоты на 21-е сутки на среде (МС контроль) — 12,3; (МС-1) — 12,8; (МС-2) — 13,1; (МС-3) — 13,0.

При микроклональном размножении раннего сорта картофеля «Алена» на 7-е сутки среднее значение высоты на среде (МС контроль) — 4,7; (МС-1) —

5,0; (МС-2) — 5,5; (МС-3) — 5,9; среде на 14-е сутки среднее значение высоты на среде (МС контроль) — 6,6; (МС-1) — 7,2; (МС-2) — 7,4; (МС-3) — 7,3; среднее значение высоты на 21-е сутки на среде (МС контроль) — 12,0; (МС-1) — 12,7; (МС-2) — 12,9; (МС-3) — 13,1.

Данные прогнозирования динамики ростовых процессов картофеля в условиях *in vitro* при использовании различных вариантов питательных сред на основе динамических моделей обработки рядов представлены на рис. 2–4.

Показатели средней длины корней у сорта «Хозяюшка» на 7-е сутки на питательной среде (МС контроль) — 1,7; (МС-1) — 1,9; (МС-2) — 2,0; (МС-3) — 2,4; на 14-е сутки среднее значение высоты на среде (МС контроль) — 2,9; (МС-1) — 3,0; (МС-2) — 4,6; (МС-3) — 4,4; среднее значение высоты на 21-е сутки на среде (МС контроль) — 4,3; (МС-1) — 4,7; (МС-2) — 4,5; (МС-3) — 5,8.

Средняя длина корней сорта «Ермак» на 7-е сутки на питательной среде (МС контроль) — 1,9; (МС-1) — 2,0; (МС-2) — 2,4; (МС-3) — 2,9; на 14-е сутки среднее значение высоты на среде (МС контроль) — 3,7; (МС-1) — 4,8; (МС-2) — 5,9; (МС-3) — 7,0; среднее значение высоты на 21-е сутки на среде (МС контроль) — 5,4; (МС-1) — 5,4; (МС-2) — 6,6; (МС-3) — 6,4.

Показатели средней длины корней сорта «Алена» на 7-е сутки на питательной среде (МС контроль) — 1,8; (МС-1) — 1,8; (МС-2) — 2,1; (МС-3) — 2,5; на 14-е сутки среднее значение высоты на среде (МС контроль) — 3,3; (МС-1) — 4,1; (МС-2) — 4,4; (МС-3) — 4,3; среднее значение высоты на 21-е сутки на среде (МС контроль) — 5,0; (МС-1) — 5,9; (МС-2) — 5,3; (МС-3) — 5,9.

Данные прогнозирования динамики ростовых процессов корневой системы картофеля в условиях

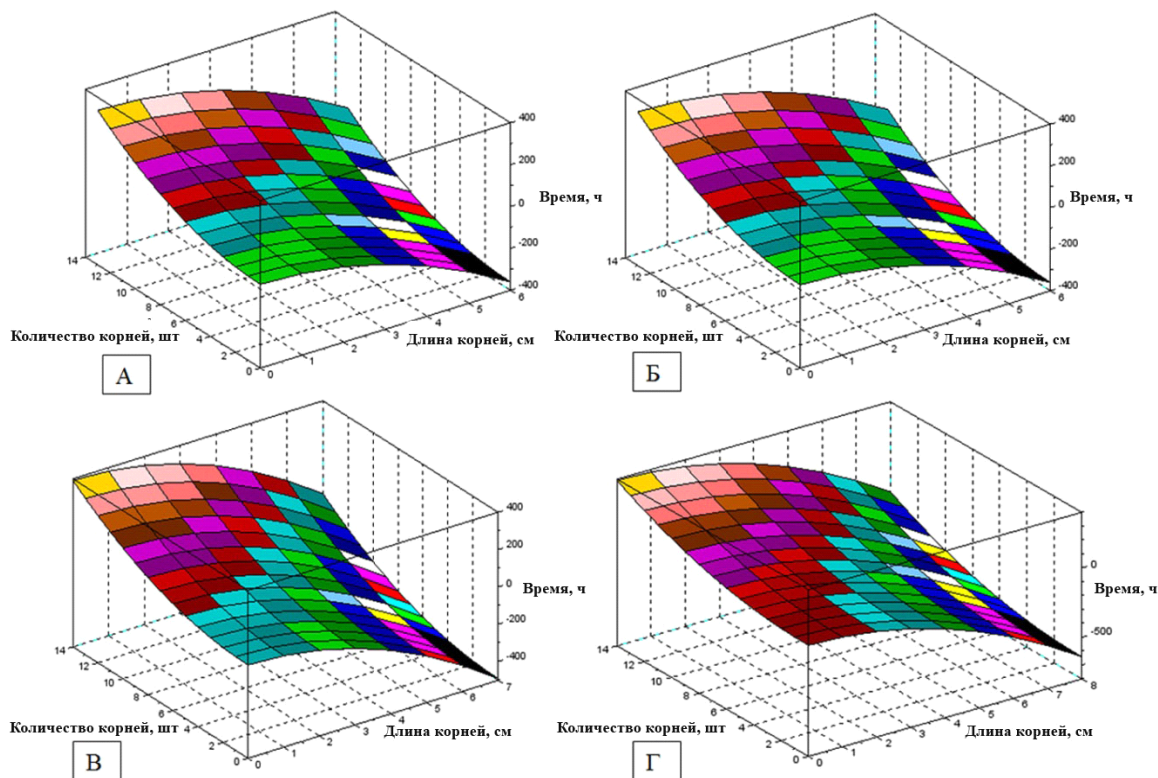


Рис. 6. Графики динамики ростовых процессов сорта картофеля «Ермак» в условиях *in vitro*: А — среда МС (контроль); Б — среда МС-1; В — среда МС-2; Г — МС-3

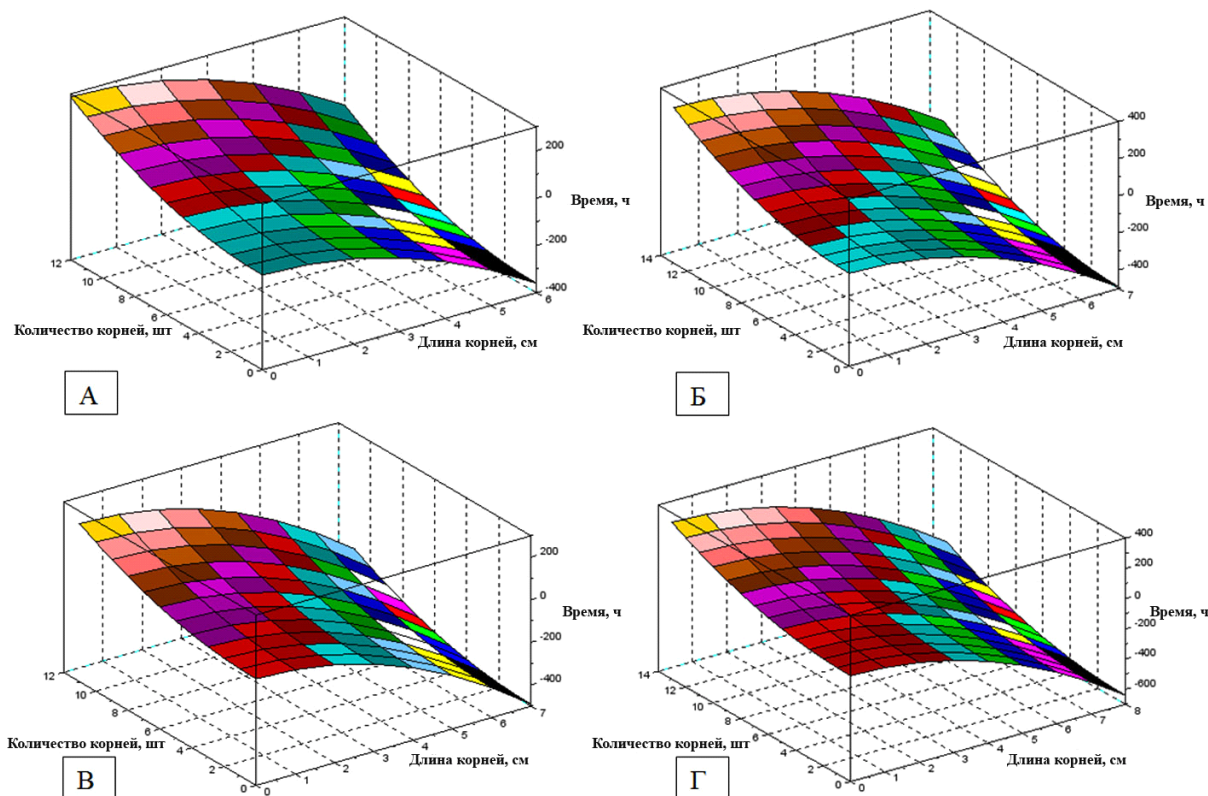


Рис. 7. Графики динамики ростовых процессов сорта картофеля «Алена» в условиях *in vitro*: А — среда МС (контроль); Б — среда МС-1; В — среда МС-2; Г — МС-3

*in vitro* при использовании различных вариантов питательных сред на основе динамических моделей обработки рядов представлены на рис. 5–7.

В результате проведенных математических расчетов моделирования ростовых процессов картофеля в условиях *in vitro*, при культивировании на разных

вариантах питательных сред на основе методов сингулярного спектрального анализа (SSA) и аппроксимации (LA), было установлено, что более лучшие ростовые процессы изучаемых сортов картофеля «Ермак», «Алена» и «Хозяюшка» наблюдалась на питательной среде МС – 3.

**Результаты исследований.** Важность научных исследований в области микрклонального размножения картофеля, а также технологий их промышленного производства, внедрения инновационных проектов согласуется с программой развития аграрного сектора страны и в комплексе позволяет объединить потенциал аграрного и научного секторов, тем самым добиться максимального экономического эффекта и прогнозирования данных.

В результате математического моделирования биотехнологического процесса микрклонального размножения картофеля в условиях *in vitro* у изучаемых сортов «Ермак», «Алена» и «Хозяюшка» более эффективные ростовые процессы наблюдались на питательной среде МС-3, содержащей в составе тиамин 1,5 мг/л, пиридоксин 1,0 мг/л, аскорбиновую кислоту 3,0 мг/л, феруловую кислоту 0,005 мг/л, кинетин 0,5 мг/л, ИУК 0,2 мг/л, сахарозу 30 000 мг/л.

#### Библиографический список

1. Wu Z., Zhang J., Chen J. Design, synthesis, antiviral bioactivity and three dimensional quantitative structure – activity relationship study of novel ferulic acid ester derivatives containing quinazoline moiety // Pest Management Science. 2017. P. 1 – 11.
2. Huang G. Y., Cui C., Wang Z. P. Synthesis and characteristics of (Hydrogenated) ferulic acid derivatives as potential antiviral agents with insecticidal activity // Chemistry Central Journal. 2013. Vol. 7, no. 1. P. 7 – 33.
3. Чурбанов Д. В., Щеглов А. Ю. Существование и единственность определения нелинейного коэффициента в задаче популяционной динамики // Тихоновские чтения: материалы науч. конф., 29 октября 2012 г. / МГУ им. М. В. Ломоносова. М.: МАКС Пресс, 2012. 60 с.
4. Hanno B., Ruth M. Modeling dynamic biological systems // Modeling Dynamic Biological Systems. Modeling Dynamic Systems. Springer, Cham. 2014. P. 3 – 28. DOI: 10.1007/978 – 3 – 319 – 05615 – 9\_1.
5. Ciarletta P., Hillen T., Othmer H. Mathematical Models and Methods for Living Systems: Levico Terme // Springer. 2016. Vol. 2167. 324 p. ISBN 3319426788.
6. Лебедева Н. В. Ускоренное размножение ранних сортов картофеля в условиях *in vitro* и его использование в семеноводстве северо-запада РФ: дис. ... канд. с.-х. наук. Великие Луки, 2015. 166 с.

7. Ковалев А. И. Совершенствование приемов оздоровления и возделывания семенного картофеля в условиях нечерноземной зоны России: дис. ... канд. с.-х. наук. Великие Луки, 2015. 122 с.

8. Лоскутов А. Ю. Анализ временных рядов. Курс лекций. URL: [http://chaos.phys.msu.ru/loskutov/PDF/Lectures\\_time\\_series\\_analysis.pdf](http://chaos.phys.msu.ru/loskutov/PDF/Lectures_time_series_analysis.pdf) (дата обращения: 20.07.2017).

9. Sanei S., Hassani H. Singular spectrum analysis of biomedical signals. CRC Press, 2015. 260 p. ISBN 9781466589278.

10. Лоскутов А. Ю. Очарование хаоса // Успехи физических наук. 2010. Т. 180, №. 12. С. 1305 – 1329.

11. Фролов А. Е., Хрущев С. С., Плюснина Т. Ю. Анализ изменений кривой индукции флуоресценции клеток микроводорослей *Chlorella SP.* в процессе роста методом мультиэкспоненциальной аппроксимации // Экологические и биологические системы: сб. ст. М.: Изд-во МГУ им. М. В. Ломоносова, 2015. С. 198 – 200.

12. Ruelle D., Takens F. On the nature of turbulence // Communications in Mathematical Physics. 1971. Vol. 20, Issue 3. P. 167 – 192. DOI: 10.1007/BF01646553.

13. Grassberger P., Procaccia I. Measuring the strangeness of strange attractors // Physica D. 1983. Vol. 9. P. 189 – 208.

14. Хорев П. Объектно-ориентированное программирование с примерами на C#. М.: Форум, НИЦ ИНФРА-М, 2016. 200 с. ISBN 978-5-00091-144-0.

**КИРГИЗОВА Ирина Васильевна**, аспирантка кафедры «Химическая технология и биотехнология» Омского государственного технического университета.

Адрес для переписки: [irina.kz-89@mail.ru](mailto:irina.kz-89@mail.ru)

**ГАДЖИМУРАДОВА Айсарат Махмудовна**, PhD-докторант по направлению 6D060700 «Биология» кафедры «Биотехнология и микробиология» РГП на ПХВ «Евразийского национального университета им. Л. Н. Гумилева».

Адрес для переписки: [aisarat3878@mail.ru](mailto:aisarat3878@mail.ru)

**КАЛИЕВ Нурлан Берикович**, магистрант гр. ВМКМ-22 кафедры «Вычислительная математика и кибернетика» Казахстанского филиала МГУ им. М. В. Ломоносова.

Адрес для переписки: [nkaliev1@gmail.com](mailto:nkaliev1@gmail.com)

Статья поступила в редакцию 05.09.2017 г.

© И. В. Киргизова, А. М. Гаджимурадова, Н. Б. Калиев

## Книжная полка

**Нестеров, С. А. Информационная безопасность : учеб. и практ. / С. А. Нестеров. – М. : Юрайт, 2017. – 322 с. – ISBN 978-5-534-00258-4.**

Серия «Университеты России» позволит высшим учебным заведениям нашей страны использовать в образовательном процессе учебники и учебные пособия по различным дисциплинам, подготовленные преподавателями лучших университетов России и впервые опубликованные в издательствах университетов. Все представленные в этой серии учебники прошли экспертную оценку учебно-методического отдела издательства и публикуются в оригинальной редакции. В учебнике «Информационная безопасность» автора Нестерова С. А. хорошо представлены защиты информации, основы криптографии, рассмотрены соответствующие описательные примеры, представлены методы расчета и статистические данные. На данный момент ряд данных (ГОСТы) информационной безопасности, приводимый в учебнике, устарел. Однако это существенно не влияет на процесс обучения анализа и управления рисками в сфере информационной безопасности, и произошедшие изменения при необходимости могут быть учтены преподавателями. Пособие может использоваться в системах повышения квалификации в рамках образовательной программы дополнительного профессионального образования «Информатика и вычислительная техника». Также может быть полезно широкому кругу специалистов в области информационных технологий.



Уфимский государственный  
авиационный технический  
университет, г. Уфа

Омский государственный  
технический университет,  
г. Омск

Омский государственный  
университет путей сообщения,  
г. Омск

## **ОБ ОЦЕНКЕ ВОЗМОЖНОСТЕЙ ЧЕЛОВЕКА ПО РАСПОЗНАВАНИЮ РУКОПИСНЫХ ОБРАЗОВ В ПРОЦЕССЕ ИХ ВОСПРОИЗВЕДЕНИЯ НА ЭКРАНЕ МОНИТОРА**

**Проведен эксперимент по идентификации авторства рукописных контрольных слов людьми. Процент ошибок в зависимости от количества одновременно идентифицируемых подписантов составил от 2,6 % до 17,6 %. По результатам эксперимента искусственный интеллект существенно превосходит естественный по надежности распознавания подписантов по рукописным образам. Установлено, что естественный интеллект использует для распознавания рукописных динамических графических образов признаки, не зависящие от скорости их воспроизведения.**

**Ключевые слова:** идентификация подписантов, естественный интеллект, искусственный интеллект, распознавание образов, особенности воспроизведения подписи.

**Работа выполнена при финансовой поддержке РФФИ (грант № 16-07-01204).**

Идентификацией человека по особенностям написания подписи занимается графология, наука, изучающая зависимости между почерком и личностью человека, его характером. С достигнутыми в этом направлении результатами можно ознакомиться в [1–8]. Графологическая экспертиза применяется в бизнесе, медицине, криминалистике, спорте, педагогике и в других областях. Осуществление почерковедческой экспертизы человеком — дорогостоящая и длительная операция, которая по некоторым обобщенным оценкам дает в среднем одну ошибку на тысячу случаев (экспертиз) [9]. При принятии решения специалист-графолог не только исследует рукописный образ, но и учитывает дополнительную информацию о подписанте для глубинного анализа, затрагивающего особенности его личности.

На сегодняшний день наблюдается рост числа научных работ, посвященных разработке методов автоматической идентификации и верификации личности человека по подписи [10–17]. При идентификации подписанта его автограф сравнивается с эталонами автографа других (нескольких) подписантов (сравнение один ко многим), при верифика-

ции подписи ее образец сравнивается с эталоном автографа конкретного субъекта (сравнение один к одному). Вместе с тем данные методы не способны учитывать дополнительную информацию о подписанте и анализируют только доступные признаки — особенности внешнего вида подписи, параметры изменения скорости и давления пера при ее написании. Достигнутая надежность данных процедур составляет порядка 0,5–1 % ошибок в режиме верификации [10, 12]. При идентификации субъектов вероятность ошибок возрастает в зависимости от количества идентифицируемых подписантов. Например, в работе [18] при идентификации 20 субъектов по подписи средний процент ошибок составил 1,2 %, в работе [19] при идентификации 80 подписантов получено в среднем 1,75 % ошибок. На первый взгляд, получается, что надежность распознавания подписи человеком примерно в 5–10 раз выше, чем надежность решений, принимаемых искусственным интеллектом. Однако такое сравнение является некорректным, по крайней мере, по двум причинам:

1. Человеку-эксперту доступна дополнительная информация о подписанте, которой нет в распоряжении у распознающего автомата.

2. Профессиональный графолог при проведении экспертизы может использовать имеющиеся средства идентификации, верификации или классификации рукописных образов, т.е. прибегать к помощи искусственного интеллекта при принятии собственного решения.

В силу изложенного выше возникает ряд вопросов. Какова вероятность ошибки верификации подписи естественным интеллектом? Какие признаки использует человек при распознавании рукописного образа? Чьи решения являются более надежными — человека или машины? Анализ научной литературы не дает четких ответов на данные вопросы. Поэтому определить вероятность ошибки идентификации человека по его подписи (или другому слову) естественным интеллектом можно только на основании эксперимента.

Настоящая статья описывает эксперимент, цель которого установить природные возможности естественного интеллекта по распознаванию воспроизводимых рукописных образов (далее подписей).

В литературе имеется много информации о признаках подписи и почерка в целом, которые характеризуют личность, характер, а также психофизиологическое состояние человека [1–8]: направление, длина, особенности написания начала и конца подписи (либо другого рукописного образа), размеры, закругленность и острота, связанность и разорванность букв, сила нажатия (давление) при написании букв, наличие вертикальных линий в подписи, скорость написания рукописного образа (всего известно более 25 характеристик). Различные характеристики почерка дают в совокупности ценную информацию о темпераменте человека, его характере, состоянии в момент написания, отношении к предмету и содержанию написанного. Психологический анализ почерка предполагает учет всех известных сведений о данном индивиде и обстоятельствах написания анализируемого текста. Результаты современных исследований подтверждают имеющуюся взаимосвязь почерка с некоторыми психическими особенностями человека. По мнению многих психологов, в почерке больше, в чем в любой другой двигательной активности человека, проявляется его психическая сущность. Неосознанное напряжение определенных групп мышц сопровождается эмоциональные переживания. Различия в тоне разных мышечных групп будут источником различий и в расположении линий на бумаге. Некоторые геометрические понятия символичны для людей. Например, «острые углы», как правило, ассоциируются с упорством, резкостью, неуступчивостью, что зачастую находит свое отражение в написании букв: но если упорство и резкость человеку неприятны, он будет избегать в письме острых углов, потому что знает об этой ассоциации.

При машинном распознавании могут анализироваться давление и скорость письма, а также характеристики внешнего вида рукописных образов. Однако в этом случае используется несколько иной подход для выделения признаков — конкретных параметров (физических величин), которые характеризуют подписанта. Если эксперт-графолог оценивает наличие или степень выраженности отдельных элементов письма (дифференцирует рукописный образ), то средство автоматической идентификации человека вычисляет интегральные величины, используемые в дальнейшем для обучения и распознавания: коэффициенты ряда Фурье, вейвлет-разложений, коэффициенты корреляции между функциями коор-

динат и давления пера на устройстве ввода, вычисление матрицы расстояний между равноудаленными точками подписи [10, 11].

При сравнении способностей естественного интеллекта с искусственным необходимо воссоздать максимально равные условия. Прежде всего, нужно выбрать фрагменты почерка для распознавания. Использование в качестве распознаваемого слова автографа видится нецелесообразным, т.к. в этом случае естественный интеллект будет «цепляться» только за графический образ подписи, не анализируя другие признаки — динамику воспроизведения рукописного слова. Искусственный интеллект работает иным образом. Алгоритмы, лежащие в основе распознающего автомата, анализируют числа, а не абстрактно визуальные образы, как человек, автомат не способен к субъективизму оценочных суждений (скопировать модель работы мышления человека пока никому не удалось). Поэтому в рамках эксперимента решено использовать фиксированное слово, которое будет воспроизводиться разными людьми на графическом планшете.

Количество информации, которое смогут использовать человек и автомат для того, чтобы отнести реализацию рукописного слова к одному из знаковых образов, должно быть сравнимым (либо идентичным). В процессе принятия решений естественным интеллектом учитываться должно не только изображение слова, но и динамика его написания, т.е. нужно, чтобы испытуемый наблюдал за тем, как слово воспроизводится на экране в том же темпе, в котором его написал субъект. Знания о динамике давления можно передать за счет использования градаций серого цвета (чем сильнее нажатие, тем темнее изображение на локальном участке рукописного образа).

Также требуется обеспечить сравнимый объем обучающей выборки.

На основе описанных принципов был разработан план эксперимента. На первом этапе осуществлялось формирование исходного материала для проведения исследований. Для этой цели было разработано программное обеспечение, позволяющее воспроизводить рукописные слова на графическом планшете. Десять человек по 50 раз каждый написали пароль «Безопасность» на графическом планшете при помощи разработанного программного обеспечения (всего получено 500 реализаций рукописного образа). В качестве устройства ввода использовался графический планшет Wacom Intuos. В ПЭВМ были занесены координаты точек пера  $x(t_i)$ ,  $y(t_i)$  и давления  $p(t_i)$ , где  $i$  — номер момента времени  $t$ , частота моментов времени зависит от частоты дискретизации графического планшета и в рамках эксперимента составляла 6 миллисекунд.

На втором этапе производилось обучение испытуемых, которые в дальнейшем участвовали в распознавании субъектов по рукописному паролю. В качестве испытуемых были привлечены молодые люди в возрасте от 18 до 25 лет, мужчины и женщины в равном соотношении, число которых составляло 50 человек. На экране монитора одна за другой по имеющемуся в ПЭВМ данным формировались 20 произвольных реализаций рукописного слова определенного человека в темпе их получения на графическом планшете (такое количество обучающих примеров требует ГОСТ Р 52633.5-2011 для настройки нейросетевых преобразователей биометрия-код, сравнимое число образцов использовалось в приведенных работах [10, 11, 18, 19]). Испыту-

Вероятности ошибок верификации рукописных паролей  
естественным интеллектом при различной скорости их воспроизведения

Скорость воспроизведения (время между прорисовкой соседних точек подписи), мс	Ошибка 2-го рода (ложного опознавания неизвестной подписи как известной)	Ошибка 1-го рода (ложного не опознавания известной подписи)
3 (скорый темп)	0,003	0,1
6 (нормальный темп)	0,02	0,08
12 (замедленный темп)	0,012	0,09

емые запоминают особенности предъявляемого образа, чтобы в будущем отличить данный образ от образов рукописных слов других людей.

Третий этап — распознавание субъектов по предъявляемым реализациям рукописного слова-пароля. На экране в случайном порядке и в сходном темпе формируется одна из оставшихся 300 подписей. Испытуемые, наблюдая за её воспроизведением, каждый независимо от других должен принимать решение, кому принадлежит представленная подпись. После предъявления всех 300 подписей определяется: сколько раз принято (каждым сотрудником) верное решение и сколько раз ошибочное. Для фиксирования решений каждому испытуемому был дан специальный бланк.

Этап 4 — эксперимент повторяется еще 2 раза, но при условии, что скорость воспроизведения подписей в одном случае превышает зафиксированную на стадии их формирования в два раза, в другом случае — она медленнее в два раза. Результаты 3-го и 4-го этапов приведены в табл. 1 и на рис. 1.

Этап 5 — распознавание субъектов в режиме идентификации. Обучение испытуемых проводилось по подписям 2-х лиц (из 10). Распознавание осуществлялось как при обычной скорости воспроизведения, так и в ускоренном и замедленном темпах. Результаты этапа 5 представлены в табл. 2 ( $Q_i$  — вероятность ошибочной классификации распознаваемого субъекта как субъекта под номером  $i$ ;  $Q_u$  — вероятность ошибочной классификации распознаваемого субъекта как неизвестного) и на рис. 2.

Этап 6. Далее эксперимент повторялся, но количество распознаваемых лиц возрастало (обучение производилось по реализациям рукописных паролей 3-х лиц, 4-х лиц и т.д.). Было принято решение повышать число идентифицируемых испытуемых субъектов до тех пор, пока вероятности ошибок распознавания не примут высокие значения, при которых дальнейшее увеличение количества классифицируемых образов было бы бесполезным. Результаты этапа 6 представлены в табл. 3–5.

Производилась оценка вероятностей правильной идентификации каждого отдельного лица, вероятность признания злоумышленников (неизвестный образ принимался как один из известных), вероятность ошибки отнесения одного известного лица к другому, вероятность ложного не признания известного лица (признания его как «чужого»).

Описанный эксперимент был повторен с использованием контрольных слов «Авторизация», «Идентификация», «Экранирование». В представленных материалах (рис. 1–3, табл. 1–5) приведены обобщенные оценки.

Как видно из табл. 1 и 2, изменение скорости воспроизведения подписей на экране приводит к незначительному увеличению количества ошибок пропуска цели и снижению вероятности ошибки лож-

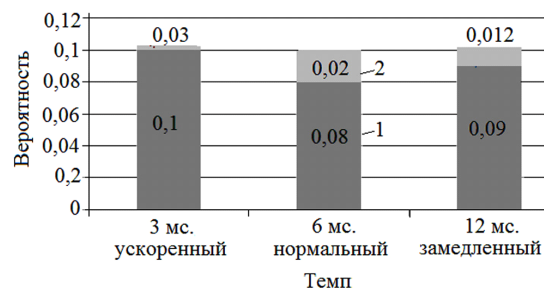


Рис. 1. Оценки вероятностей ошибок распознавания субъектов в режиме верификации при различной скорости воспроизведения рукописных паролей: 1 — ошибка 1-го рода, 2 — ошибка 2-го рода

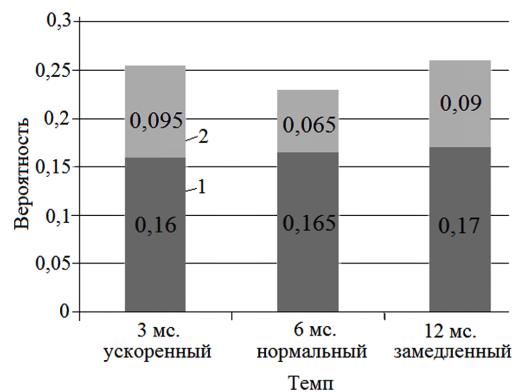


Рис. 2. Оценки вероятностей ошибок идентификации 2-х субъектов при различной скорости воспроизведения рукописных паролей: 1 — ошибка 1-го рода, 2 — ошибка 2-го рода

ного распознавания. При этом средняя вероятность ошибки распознавания почти не изменяется. Таким образом, по приведенным результатам можно заключить, что скорость воспроизведения не оказывает существенного влияния на надежность принимаемых естественным интеллектом решений (рис. 1 и 2). Данный результат можно отнести к категории неожиданных: естественный интеллект использует признаки, не зависящие от скорости воспроизведения рукописного пароля.

Результаты проведенного эксперимента указывают на то, что в режиме идентификации естественный интеллект человека существенно уступает искусственному по надежности распознавания. Видимо, оценка надежности графологической экспертизы в работе [9] дана с учетом использования специальных профессиональных техник и средств автоматического анализа подписей с привлечением дополнительных знаний о подписанте (тип темперамента,

**Таблица 2**  
**Вероятности ошибок идентификации естественным интеллектом**  
**2-х субъектов по рукописному паролю**  
**при различной скорости их воспроизведения**

Обычный темп (6 мс.)			
Распознаваемый субъект	$Q_1$	$Q_2$	$Q_u$
Субъект 1		0,02	0,21
Субъект 2	0		0,1
Неизвестный субъект	0,043	0,022	
Ускоренный темп (3 мс.)			
Распознаваемый субъект	$Q_1$	$Q_2$	$Q_u$
Субъект 1		0,03	0,17
Субъект 2	0		0,12
Неизвестный субъект	0,075	0,02	
Замедленный темп (12 мс.)			
Распознаваемый субъект	$Q_1$	$Q_2$	$Q_u$
Субъект 1		0,03	0,16
Субъект 2	0		0,15
Неизвестный субъект	0,07	0,02	

**Таблица 3**  
**Вероятности ошибок идентификации естественным интеллектом 3-х субъектов**  
**по рукописному паролю при различной скорости их воспроизведения**

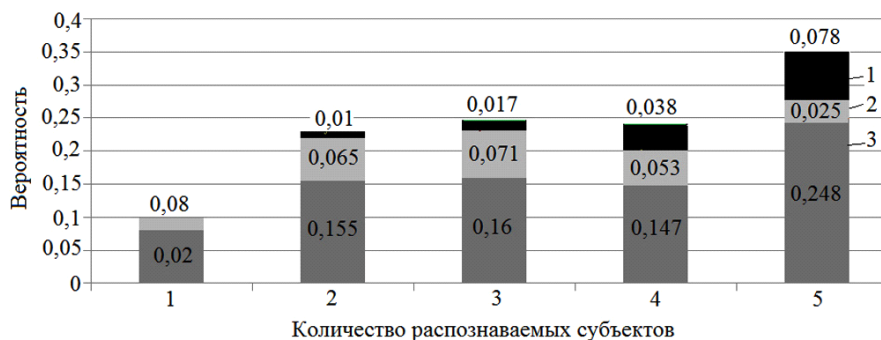
Распознаваемый субъект	$Q_1$	$Q_2$	$Q_3$	$Q_u$
Субъект 1		0,03	0	0,12
Субъект 2	0		0,01	0,16
Субъект 3	0	0,01		0,2
Неизвестный субъект	0,017	0,022	0,032	

**Таблица 4**  
**Вероятности ошибок идентификации естественным интеллектом 4-х субъектов**  
**по рукописному паролю при различной скорости их воспроизведения**

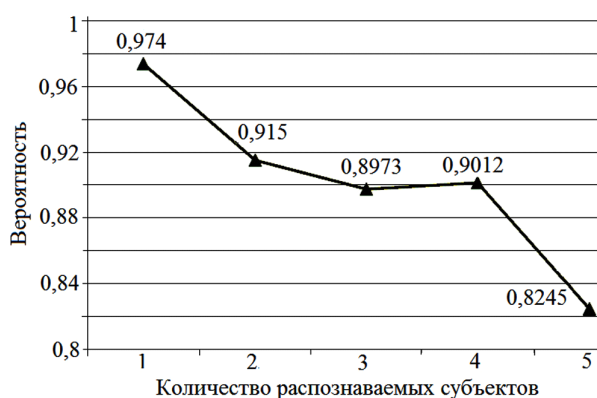
Распознаваемый субъект	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$
Субъект 1		0,01	0,01	0,02	0,11
Субъект 2	0		0,05	0,02	0,12
Субъект 3	0	0,02		0	0,27
Субъект 4	0,01	0,01	0		0,09
Неизвестный субъект	0,01	0,023	0,017	0,003	

**Таблица 5**  
**Вероятности ошибок идентификации естественным интеллектом**  
**5-ти субъектов по рукописному паролю при различной скорости их воспроизведения**

Распознаваемый субъект	$Q_1$	$Q_2$	$Q_3$	$Q_4$	$Q_5$	$Q_u$
Субъект 1		0,05	0,04	0,02	0	0,07
Субъект 2	0		0	0,02	0,02	0,09
Субъект 3	0	0,03		0,04	0,02	0,4
Субъект 4	0	0,02	0,03		0,01	0,35
Субъект 5	0	0,02	0	0,07		0,33
Неизвестный субъект	0,005	0	0,015	0	0,005	



**Рис. 3. Изменение вероятностей ошибок идентификации в зависимости от количества известных субъектов:**  
 1 — ошибка ложного опознавания,  
 2 — ошибка ложного неопознавания;  
 3 — ошибочная идентификация известных субъектов



**Рис. 4. Изменение вероятности правильной идентификации в зависимости от количества известных субъектов**

возраст, пол и т.д.). При незначительном увеличении количества идентифицируемых субъектов число ошибочных решений, сделанных человеком, стремительно возрастает (вероятность верного распознавания снижается, рис. 4). Видимо, это обусловлено ограничениями кратковременной памяти человека. В режиме верификации естественный интеллект также уступает искусственному, но не столь значительно (с учетом погрешностей).

При увеличении количества образов субъектов вероятность ложного опознавания неизвестного субъекта как одного из известных снижается, в то время как вероятность других ошибок возрастает.

Проведен эксперимент по распознаванию подписантов по динамике написания паролей людьми (количество испытуемых составило 50). Получены следующие результаты по сравнению возможностей естественного интеллекта по распознаванию подписей с методами автоматического распознавания субъектов по подписи [10, 12, 18, 19] (с использованием алгоритма последовательного применения формулы гипотез Байеса [12, 18, 19], искусственных нейронных сетей [10] и сетей квадратичных форм [10]):

— естественный интеллект используют для распознавания рукописных динамических графических образов; признаки, не зависящие от скорости их воспроизведения;

— методы автоматического распознавания подписей (в частности, из работ [10, 12]) превосходят естественный интеллект при верификации рукописных образов, в режиме верификации средняя

вероятность ошибки распознавания рукописного слова человеком составила 0,026;

— методы автоматического распознавания подписей (в частности, из работ [18, 19]) существенно превосходят естественный интеллект при идентификации рукописных образов, человек не способен запомнить множество различных образцов подписей, средняя вероятность ошибки идентификации 2, 3, 4 и 5 рукописных слов человеком составила 0,085, 0,103, 0,099, 0,176.

#### Библиографический список

1. Кравченко В. И. Графология: характер по почерку. СПб.: ГУАП, 2006. 92 с. ISBN 5-0880-189-3.
2. Манцетова А. И. Теоретические (естественнонаучные) основы судебного почерковедения. 2-е изд. М.: Наука, 2006. 442 с. ISBN 5-02-035816-9.
3. Потемкина О. Ф. Психологический анализ рисунка и текста. СПб.: Речь, 2006. 524 с. ISBN 5-9268-0362-4.
4. Судилова Е. В. Разгадываем почерк. Графология: моногр. СПб.: Питер, 2007. 159 с. ISBN 978-5-91180-364-3.
5. Лабо М. Г. О чем говорит почерк? Практическая графология. Записки полковника ФСБ. СПб.: ДНК, 2007. 128 с. ISBN 978-5-901562-70-3.
6. Гольдберг И. Психология почерка. М.: АСТ МОСКВА, 2008. 161 с. ISBN 978-5-9757-0376-7, 978-5-9713-9270-5.
7. Гольдберг И. Почерк и типы личности. М.: АСТ МОСКВА, 2008. 192 с. ISBN 978-5-9757-0324-8, 978-5-9713-8392-5.
8. Норман Д. О чем говорит почерк? Графология для начинающих. СПб.: Вектор, 2010. 140 с. ISBN 978-5-9684-1450-2.
9. Болл Р. М., Коннел Дж. Х., Панканти Ш. [и др.]. Руководство по биометрии. М.: Техносфера, 2007. 370 с. ISBN 978-5-94836-10-93.
10. Сулавко А. Е., Еременко А. В., Толкачева Е. В. [и др.]. Комплексирование независимых биометрических признаков при распознавании субъектов на основе сетей квадратичных форм, перцептронов и меры ХИ-модуль // Информационно-управляющие системы. 2017. № 1 (86). С. 50 – 62. DOI: 10.15217/issnl684-8853.2017.1.50.
11. Lozhnikov P. S., Sulavko A. E., Eremenko A. V., Volkov D. A. Methods of Generating Key Sequences based on Parameters of Handwritten Passwords and Signatures // Information. 2016. № 7. P. 59.
12. Daramola S. A., Ibiyemi T. S. Efficient Online Signature Verification System // International Journal of Engineering and Technology IJET-IJENS. 2010. Vol. 10, no. 4. P. 42 – 46.
13. Vahab Iranmanesh, Sharifah Mumtazah Syed Ahmad, Wan Azizun Wan Adnan, Salman Yussof, Olasimbo Ayodeji Arigbabu, Fahad Layth Malallah. Online Handwritten Signature Verification Using Neural Network Classifier Based on Principal Component

Analysis // The Scientific World Journal. 2014. Vol. 2014. P. 1–8. DOI: 10.1155/2014/381469.

14. Hao Chang, Huaizhong Bao, Yutao Sun, Sulin Wei. Online Signature Verification Based on Feature Combination and Classifier Fusion // Journal of Information & Computational Science. 2013. Vol. 10, no. 6. P. 1613–1621. DOI: 10.1504/IJCS.2014.060713.

15. Nan Xu, Li Cheng, Yan Guo, Xiaogang Wu and Jiali Zhao. Online handwritten signature verification by using back propagation neural networks // Communication Software and Networks (ICCSN), 2011 IEEE 3rd Intern. Conf. on, May 27–29, 2011. DOI: 10.1109/ICCSN.2011.6013611.

16. Нигрей А. А. Идентификация человека по подписи. Роль технических наук в развитии общества: сб. тр. II Междунар. науч.-практ. конф. Кемерово, 2017. С. 236–239.

17. Нигрей А. А. Исследование изменения динамики подписи подписанта на графическом планшете при изменении его психофизиологического состояния. Безопасность городской среды: материалы IV Междунар. науч.-практ. конф. Омск: Изд-во ОмГТУ, 2017. С. 383–385.

18. Ложников П. С., Сулавко А. Е., Самотуга А. Е. Модель защиты гибридных документов на основе рукописных подписей их владельцев с учетом психофизиологического состояния подписантов // Вопросы защиты информации. 2016. № 4. С. 47–59.

19. Еременко А. В. Повышение надежности идентификации пользователей компьютерных систем по динамике написания паролей: автореф. дис. ... канд. техн. наук. Омск, 2011. 20 с.

**ВАСИЛЬЕВ Владимир Иванович**, доктор технических наук, профессор (Россия), заведующий кафедрой «Вычислительная техника и защита информации» Уфимского государственного авиационного технического университета.

**СУЛАВКО Алексей Евгеньевич**, кандидат технических наук, старший преподаватель кафедры «Комплексная защита информации» Омского государственного технического университета (ОмГТУ).

**ЖУМАЖАНОВА Самал Сагидуллоевна**, аспирант кафедры «Комплексная защита информации» ОмГТУ.

**НИГРЕЙ Алексей Андреевич**, аспирант кафедры «Информационная безопасность» Омского государственного университета путей сообщения.

Адрес для переписки: nigrey.n@mail.ru

Статья поступила в редакцию 27.07.2017 г.

© В. И. Васильев, А. Е. Сулавко, С. С. Жумажанова, А. А. Нигрей

## Книжная полка

**Информатика для экономистов : учеб. / Под ред. В. П. Полякова. – М. : Юрайт, 2017. – 524 с. – ISBN 978-5-9916-4367-2.**

Рассматриваются основные аспекты процессов информатизации и формирования информационного общества, теоретические основы, понятия и определения информатики и экономической информатики, техническая и программная база информационных систем и технологий, базовые информационные технологии для решения экономических и управленческих задач. Соответствует Федеральному государственному образовательному стандарту высшего образования четвертого поколения. Учебник предназначен для студентов, обучающихся по направлениям подготовки бакалавров экономики и менеджмента, а также аспирантов и преподавателей экономических вузов. Может быть полезен для магистрантов, обучающихся по направлениям «магистр экономики» и «магистр менеджмента» для актуализации знаний, умений и навыков по применению информационных и коммуникационных технологий для анализа данных и принятия управленческих решений.

**Антонов, А. В. Системный анализ : учеб. /А. В. Антонов. – 4-е изд., перераб. и доп. – М. : Инфра-М, 2017. – 368 с. – ISBN 978-5-16-011865-9.**

В учебнике изложены методологические вопросы системного анализа. Описаны этапы и процедуры проведения системных исследований, сформулированы цели и задачи системного анализа. Большое внимание уделено вопросам построения моделей сложных систем, проверке их адекватности, процедурам их формирования, методам оценки параметров. Рассмотрены математические методы и модели системного анализа, постановки задач, описаны области их применения. Изложены численные методы решения типовых задач системного анализа. Приведены методы выбора и принятия решений. Содержание учебника соответствует требованиям Федерального государственного образовательного стандарта высшего образования последнего поколения. Для студентов, обучающихся по направлению 09.03.01 «Информатика и вычислительная техника», а также для аспирантов и инженеров.

**Бессмертный, И. А. Интеллектуальные системы : учеб. / И. А. Бессмертный, А. Б. Нугуманова, А. В. Платонов. – М. : Юрайт, 2017. – 244 с. – ISBN 978-5-534-01042-8.**

В учебнике рассматриваются основы построения интеллектуальных систем и методов извлечения знаний. Наряду с традиционными интеллектуальными системами, базирующимися на продукционной модели знаний, в учебнике уделяется внимание также методам поиска информации в неформализованных и слабо формализованных источниках, в частности методам машинного обучения и статистическим методам обработки естественно-языковых текстов. Содержание учебника соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования. Для студентов высших учебных заведений, аспирантов, преподавателей, слушателей курсов повышения квалификации, а также разработчиков программного обеспечения.

## ПРОБЛЕМА ИДЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЯ КОМПЬЮТЕРНОЙ СИСТЕМЫ ПО ТРЕХМЕРНЫМ ИЗОБРАЖЕНИЯМ ЛИЦА

В данной работе рассматривается актуальная проблема защиты информационных ресурсов посредством проведения скрытой идентификации личности субъекта, выполняющего должностные обязанности на рабочем месте, по трехмерному изображению лица. Приводится краткое описание существующих и разрабатываемых технологий 3D-съемки, реализованных в камерах, входящих в стандартное оснащение компьютеров и ноутбуков. Обозначены достигнутые результаты в данной области, описан ряд используемых идентификационных признаков лица.

**Ключевые слова:** анализ, безопасность, биометрия, идентификация, особенности лиц, распознавание.

*Работа выполнена при финансовой поддержке РФФИ (грант №16-07-01204).*

**Введение.** Согласно аналитическим отчетам консалтинговых компаний за последние годы неизменным остается следующий вывод: более половины всех финансовых потерь от преступлений в компьютерной сфере связаны с деятельностью собственных сотрудников компаний. В 2016 году количество инцидентов по вине инсайдеров перешагнуло рубеж в 1300. Показательными являются следующие цифры: в 36 % случаев виновниками утечек информации были сотрудники компаний [1]. Объем ущерба, наносимого мировой экономике, только от этих преступлений в последние годы колеблется в районе 10 млрд \$ [2].

Анализ ситуации показывает, что причины большинства инцидентов кроются в превышении должностных полномочий авторизованными лицами, ошибками в работе и неправомерном использовании идентификаторов сотрудников организации. Проблема идентификации пользователей для разграничения доступа к информационным ресурсам решается уже на протяжении десятилетий. Однако существующие подходы к идентификации не учитывают того факта, что авторизованный пользователь сам может являться злоумышленником либо от его имени может действовать другое лицо.

Авторам работы видится перспективным использование технологий получения 3D-моделей лица для непрерывной идентификации личности в скрытом режиме при условии решения ряда проблем: миниатюризации камер, ослабления ограничений по

условиям съемки субъектов доступа, снижение стоимости оборудования для 3D-съемки.

Целью данной работы является определение достигнутых результатов в области распознавания человека по 3D-изображениям лица при помощи малогабаритных камер.

**Оборудование, позволяющее получить трехмерное изображение лица.** Существует множество методов распознавания по геометрии лица и головы. Все они основаны на том, что черты лица и форма черепа каждого человека индивидуальны. Для построения высокоточных трехмерных моделей головы и черепа требуется осуществлять съемку из нескольких точек и сформировать модель, учитывая произвольные движения, ракурс, поворот головы субъекта относительно камер. Такие условия на практике выполнимы, если для идентификации личности использовать миниатюрную 3D-видеокамеру, располагающуюся в корпусе мобильного устройства или персонального компьютера, и проводить идентификацию оператора во время работы с устройством. До недавнего времени не существовало подходящих типов устройств, позволяющих реализовать эту идею.

Большинство существующих систем идентификации по 3D-изображению лица используют лазерное сканирование [3], стереокамеры [4] или определение структуры объекта по отображению движения [5] для получения 3D-моделей лица. Стоимость сканеров для получения трехмерных изображений

Таблица 1  
Достигнутая надежность распознавания субъектов  
по трехмерному изображению лица

Метод	FRR, % при FAR = 0,1%
Kakadiaris et al. [12] (2007)	3
Faltemier et al. [13] (2008)	2,8
Alyüz et al. [14] (2009)	2,5
Queirolo et al. [15] (2010)	1,6
Spreeuwiers [16] (2011)	1

лиц достаточно высока (более 500 тыс. руб.). К наиболее доступным в настоящее время по стоимости типам сенсоров-дальномеров относятся следующие камеры [6]:

1) стереокамеры. Позволяют получить карту глубины, используя две обычные камеры. Преимущества: низкая стоимость, широко распространены. Недостатки: требуются дополнительные вычислительные ресурсы для восстановления карты глубины по изображениям, полученным с помощью камер; высокая зависимость от освещения;

2) стереокамера со структурированной инфракрасной подсветкой. Преимущества: высокая точность, низкая стоимость, нетребовательна к освещенности. Недостатки: сложно найти в продаже, низкая точность на расстояниях более 2 метров;

3) TOF (Time of Flight — времяпролетная) камера, в литературе также встречается название RGB-D камера. Название RGB-D говорит о том, что камера позволяет получать помимо двумерного изображения еще и данные о глубине изображения. Преимущества: высокая точность, широко распространены, несущественная зависимость от освещенности, высокая дальность действия относительно других типов сенсоров-дальномеров. Недостатки: сравнительно высокая стоимость.

Одна из технологий получения информации о трехмерных объектах, реализованных в недорогих TOF камерах (DepthSense, Kinect, Basler) основана на использовании длинноволнового инфракрасного излучения, проецирующего свет в пространство. Затем отдельный CMOS-датчик изображения считывает ИК-свет и фиксирует его вместе с серией синхронизированных изображений. Инфракрасные паттерны, придающие глубину, декодируются чипом, разработанным компанией, чтобы создать виртуальный образ сцены или объекта.

В начале 2014 года в Лас-Вегасе, на выставке CES (Consumer Electronics Show), компанией Intel была представлена камера RealSense, позиционируемая как встраиваемое решение, которым теперь оснащаются некоторые портативные и стационарные устройства: ультрабуки, планшеты и т.д. Таким образом, 3D-камеры становятся стандартным оборудованием компьютеров.

**Достигнутые результаты в области распознавания субъектов по трехмерным изображениям лица.** Интерес к использованию более дешевого оборудования (TOF камер) для идентификации личности подчеркивается появлением множества научно-исследовательских работ на данную тематику. С использованием времяпролетных камер проведено немало исследований в области разработки человеко-машинных интерфейсов [7] и моделирования 3D-объектов [8]. В частности, работа [9], посвященная распознаванию людей по трехмерному

изображению лица, которая финансировалась по программе HPLabsInnovationResearchProgram (CW 218094) и была поддержана высшей инженерной школой и исследовательским центром коммуникационных систем EURECOM в сотрудничестве с промышленными предприятиями: ORANGE, BMW-Group, Swisscom, Cisco, SFR, STEricsson, Symantec, SAP and Monaco Telecom. Для принятия идентификационных решений в работе использовали метод расчета евклидова расстояния от эталонного облака точек в пространстве классификационных признаков до облака, полученного из предъявленного изображения. Для получения значений признаков оценивались карты глубины 3D-изображения размером 640×480 точек.

Об успехах в области идентификации личности с помощью данного вида оборудования можно судить по следующим результатам. Работа [10] посвящена разработке технологии распознавания пользователя и его поведения с помощью перцептрона, надежность составила 99 % верных классификационных решений в выборке из 30 человек. Исследование финансировалось через программу ITR&D (организатор программы MKE&KEIT).

В работе [11] получены следующие результаты для 20-ти участников эксперимента: EER = 2,1 % в случае использования пяти снимков при формировании трехмерной модели лица. База снимков составила 1054 экземпляра. Известно множество других достигнутых результатов по распознаванию человека по трехмерным моделям, наилучшие из которых представлены в табл. 1.

В основе методов распознавания человека по лицу лежат различные подходы. Среди них можно выделить точное сопоставление двух и более изображений, для этого обычно используется итеративный алгоритм поиска ближайших точек (англ. Iterative Closest Point — ICP) — алгоритм, предназначенный для сведения к минимуму разницы между двумя совокупностями точек в пространстве при восстановлении двумерных или трехмерных поверхностей из разных изображений. В работе [12] был опробован алгоритм EM-ICP (модификация ICP), реализованный для вычисления на GPU процессорах, что позволило обрабатывать облака точек в реальном режиме времени.

Другим эффективным подходом является применение широких искусственных нейронных сетей [13, 14], с использованием которых удалось достигнуть сравнительно высоких результатов, обрабатывая двумерные изображения лиц: FRR = 0,0033, FAR = 0,0061 [13], FRR = 0,0014, FAR = 0,002 [14].

**Признаки, выделяемые из трехмерных изображений лица.** Первым этапом создания системы скрытой идентификации по трехмерной модели лица является выявление устойчивых признаков. Приведем



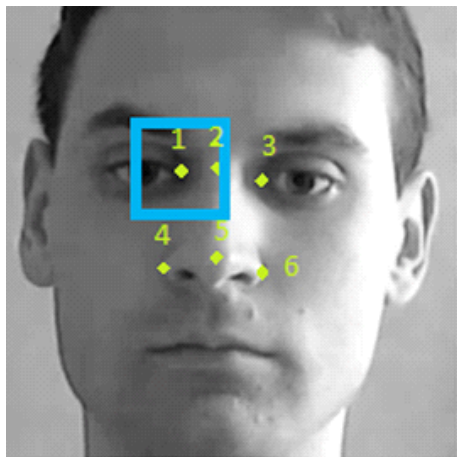


Рис. 1. Вспомогательные точки, используемые для выделения участков лица

краткую справку относительно идентификационных характеристик, встречающихся в научной литературе.

В работе [15] изображение лица разбивается на следующие области: нос, глаза, щеки и остальная часть. В [11] для указанных областей экспериментально были определены следующие веса в зависимости от их влияния на результат идентификации: 3; 2; 1; 0,5.

В качестве признаков 3D-изображения лица также используется информация о глубине изображения. В частности, в работе [16] информацию о глубине изображения (размером  $384 \times 384$ ) лица анализируют на шести участках (по  $128 \times 128$  точек), центрированных относительно ключевых (вспомогательных) точек (рис. 1). Каждый из шести участков представляют в виде векторов до и после их обработки по алгоритму DT-CWT (Dual-treecomplex-wavelettransform). Для двух векторов вычисляют матрицу перехода от одного базиса к другому.

На двумерных изображениях обычно выделение региона лица основано на использовании гистограмм цветовых диапазонов [17], зона кожи человека может быть представлена узкой полосой. Алгоритмы выде-

ления кожи и некоторых элементов лица приводятся в [18, 19]. Для выделения границ ключевых особенностей лица часто используются фильтры на основе оператора Собеля, методы Марр – Хилдрета [20], первые и вторые гауссовы производные, детектор границ Канны, оператор симметричности [21], метод Виолы – Джонса [22]. Большинство распространенных методик можно адаптировать для анализа трехмерных моделей лиц.

При идентификации лица биометрическая система автоматически выделяет и обрабатывает сведения, характеризующие отдельные (наиболее «показательные») участки и особенности: контуры носа, губ, бровей, расстояние между ними и т.п. На основе этих сведений в соответствии с общими принципами биометрических технологий формируются цифровые модели идентификаторов, которые затем сравниваются между собой. Имеется ряд факторов, негативно влияющих на точность определения особенностей лица:

- размер и ориентация лица в пространстве (возникает необходимость в нормализации изображения относительно наклона и поворота головы);

- низкое качество изображения, недостаточная яркость или контрастность;

- макияж и маскировка, эффективных способов борьбы с данными мешающими воздействиями не существует.

Предложенные для 3D-идентификации камеры позволяют получать как двумерную информацию (в виде плоского изображения лица), так и данные о глубине изображения. При анализе трехмерных изображений видится возможным использовать аналогичные признаки, детектируемые на двумерном изображении, дополненные данными о глубине:

- площади областей лба, глаз, ноздрей, губ, зрачков, бровей (рис. 2, 3), нормированных по площади области лица (нормирование используется для учета расстояния лица субъекта до камеры);

- максимальная длина и/или ширина областей губ, глаз, бровей, ушей, подбородка или челюсти (рис. 2, 3), нормированных по ширине и длине области лица, соответственно (нормирование используется для учета расстояния лица субъекта до камеры);



Рис. 2. Выделение признаков в кадре видеозаписи

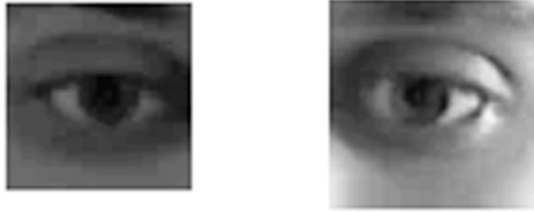


Рис. 3. Области глаз

— кратчайшие расстояния между областями ноздрей, губ, зрачков, бровей и кончика подбородка (рис. 2), нормированных по диагонали области лица (нормирование используется для учета расстояния лица субъекта до камеры);

— изменение коэффициента корреляции между участками лба, глаз, носа, губ, ушей (рис. 2) на различных кадрах видеозаписи (данная категория признаков характеризует мимику субъекта);

— коэффициент корреляции между левой и правой частями участков лица, лба, носа, губ (рис. 2), а также левым и правым участками ушей (рис. 2) и областями глаз (рис. 3) на одном кадре видеозаписи (данная категория признаков характеризует асимметрию лица субъекта);

— амплитуды гармоник с частотами от 1 до 10 Гц функций перемещения центров областей глаз, зрачков, ноздрей, бровей (рис. 2, 3) на различных кадрах видеозаписи (данная категория признаков характеризует тремор головы и отдельных частей лица субъекта, данный класс признаков не чувствителен к маскировке).

Для определения участков лица можно использовать метод Виолы – Джонса, основанный на использовании признаков Хаара. Данный метод является одним из лучших по соотношению показателей эффективности распознавания и скорости работы [23, 24]. После обработки двумерного изображения можно использовать дополнительные данные о глубине. Сегментация участков лица и их взвешивание необходимы для исключения нестабильных признаков (например, волос и ступенчатости граничных участков). Алгоритм принятия идентификационных решений может быть основан на сетях Пирсона – Хемминга, хи-модуль, а также Байеса – Хемминга, которые показывают достаточно высокие результаты в ряде задач биометрической аутентификации [13, 14].

**Методы и подходы к формированию решений при распознавании образов лица.** В настоящее время существует множество различных методов распознавания параметров лица. Ниже приведены основные, наиболее часто используемые методы:

#### 1. Гибкое сравнение на графах.

Суть данного метода сводится к эластичному сопоставлению графов, описывающих изображения лиц. Лица представлены в виде графов со взвешенными ребрами и вершинами. На этапе распознавания эталонный граф остается неизменным, в то время как другой преобразуется с целью наилучшей подгонки к первому. В подобных системах распознавания графы могут представлять собой как прямоугольную решетку, так и структуру, образованную характерными точками лица [25].

#### 2. Нейронные сети.

В настоящее время существует около десятка разновидностей нейронных сетей. Одним из самых широко используемых вариантов является сеть, по-

строенная на многослойном перцептроне, которая позволяет классифицировать поданное на вход изображение/сигнал в соответствии с предварительной настройкой/обучением сети. Эта технология позволяет распознавать изображения с высокой точностью.

Хорошие результаты в области распознавания лиц показала сверточная нейронная сеть (Convolutional Neural Network), которая является логическим развитием идей таких архитектур, как когнитрон и неокогнитрон. Если исходное изображение имеет размерность  $100 \times 100$  пикселей по трём каналам (30000 входных нейронов), а сверточный слой использует фильтры с ядром  $3 \times 3$  пикселя с выходом на 6 каналов, то тогда в процессе обучения определяется только 9 весов ядра; в таком случае данный слой требует нахождения только 162 параметров, что существенно меньше количества искомых параметров полносвязной нейронной сети [26].

#### 3. Скрытые Марковские модели.

Одним из статистических методов распознавания лиц являются скрытые Марковские модели (СММ) с дискретным временем. СММ используют статистические свойства сигналов и учитывают их пространственные характеристики. Элементами модели являются следующие параметры: множество скрытых состояний, множество наблюдаемых состояний, матрица переходных вероятностей и начальная вероятность состояний. При распознавании объекта проверяются сгенерированные для заданной базы объектов Марковские модели и ищется максимальная из наблюдаемых вероятностей того, что последовательность наблюдений для данного объекта сгенерирована соответствующей моделью [27].

#### 4. Метод главных компонент.

Главной целью данного метода является значительное уменьшение размерности пространства признаков таким образом, чтобы оно как можно лучше описывало «типичные» образы, принадлежащие большинству лиц. Используя этот метод, можно выявить различные закономерности изменчивости в обучающей выборке изображений лиц и описать эту закономерность в базисе нескольких ортогональных векторов, которые называются собственными.

Полученный один раз на обучающей выборке изображений лиц набор собственных векторов используется для кодирования всех остальных изображений лиц, которые представляются взвешенной комбинацией собственных векторов. Используя ограниченное количество собственных векторов, можно получить сжатое приближение входного изображения лица, которое затем можно хранить в базе данных в виде вектора коэффициентов, служащего ключом поиска [28].

**Заключение.** В работе исследована проблема защиты информационных ресурсов от деятельности инсайдеров посредством проведения скрытой непрерывной идентификации субъекта, выполняющего должностные обязанности за рабочим столом, по трехмерному изображению лица. На сегодняшний день 3D-камеры получили широкую распространённость и фактически входят в стандартное оснащение современных компьютеров (ноутбуков, планшетов). Анализ литературы показал, что достигнутый уровень среднего количества ошибок распознавания по 3D-изображениям стандартных камер, имеющихся в массовой продаже, находится в пределах 0,85 – 1 %. Также данные исследований говорят о том, что пространство признаков 3D-моделей лица в существующих работах в основном строится на использо-

вании информации о глубине. Данные двумерных изображений лиц могут быть комплексированы с данными о глубине изображения для снижения вероятностей ошибок, совершенных при распознавании по двумерным моделям лица. В частности, количество ошибок, допущенных в работе [13] при аутентификации субъекта по одному 2D-изображению ( $FRR = 0,0033$ ,  $FAR = 0,0061$ ) и в работе [14] при аутентификации пользователя в течение 30 секунд работы на компьютере по 2D-видеозаписи лица ( $FRR = 0,0014$ ,  $FAR = 0,002$ ), позволят повысить надежность этих решений. Использование данных о глубине также дает возможность реализовать защиту от предьявления фотографии лица.

#### Библиографический список

1. Глобальное исследование утечек конфиденциальной информации в 2016 году. Аналитический центр InfoWatch. 2016 г. URL: <https://www.infowatch.ru/analytics/reports/17479> (дата обращения: 09.08.2017).
2. Cyber attacks are costing companies millions of dollars. Business Insider. 2016. URL: <http://www.businessinsider.com/cyber-attacks-are-costing-companies-millions-of-dollars-heres-how-they-can-mitigate-those-costs-2016-2> (дата обращения: 09.08.2017).
3. Phillips P. J., Flynn P. J., Scruggs T. [et al.]. Overview of the face recognition grand challenge // IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR), 20–25 June 2005. Vol. 1. P. 947–954.
4. Medioni G., Waupotitsch R. Face modeling and recognition in 3-D // Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'03), 17 Oct. 2003. P. 232–233.
5. Medioni G., Choi J., Kuo C. H. [et al.]. Identifying noncooperative subjects at a distance using face images and inferred three-dimensional face models // IEEE Transactions on Systems, Man, and Cybernetics. Part A. Systems and Humans. Jan. 2009. Vol. 39, no. 1. P. 12–24.
6. Котюжанский Л. А. Интерфейс бесконтактного человеко-машинного взаимодействия на основе данных сенсора-дальности: дис. ... канд. техн. наук. Екатеринбург, 2014. 146 с.
7. Shotton J., Fitzgibbon A., Cook M. [et al.]. Real-time human pose recognition in parts from single depth images // IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR), 20–25 June 2011. P. 1297–1304.
8. Henry P., Krainin M., Herbst E., Ren X., Fox D. RGB-D Mapping: Using Depth Cameras for Dense 3D Modeling of Indoor Environments // In Proc. of International Symposium on Experimental Robotics (ISER). 2010. URL: <http://citeseerx.ist.psu.edu/viewdoc/download?jsessionid=07876D76952D9C4566E15E7F25A9E36F?doi=10.1.1.226.91&rep=rep1&type=pdf> (дата обращения: 12.08.2017).
9. Shotton J., Fitzgibbon A., Cook M. [et al.]. Real-time human pose recognition in parts from single depth images // IEEE Intl. Conf. on Computer Vision and Pattern Recognition (CVPR), 20–25 June 2011. 2013. P. 1297–1304. ISBN 978-1-4577-0394-2. DOI: 10.1109/CVPR.2011.5995316.
10. Choi J., Sharma A., Medioni G. Comparing Strategies for 3D Face Recognition from a 3D Sensor. 2013. 6 p. URL: <http://iris.usc.edu/outlines/papers/2013/cho-shar-med-roman13.pdf> (дата обращения: 13.08.2017).
11. Kakadiaris I. A., Passalis G., Toderici G. [et al.]. Three-dimensional face recognition in the presence of facial expressions: an annotated deformable model approach // IEEE Transactions on pattern analysis and machine intelligence. 2007. Vol. 29, Issue 4. P. 640–649. DOI: 10.1109/TPAMI.2007.1017.
12. Yonghuai Liu. Automatic registration of overlapping 3D point clouds using closest points // Image and Vision Computing. 2006. Vol. 24, no. 7. P. 762–781.
13. Сулавко А. Е., Еременко А. В., Толкачева Е. В., Борисов Р. В. Комплексирование независимых биометрических признаков при распознавании субъектов на основе сетей квадратичных форм, перцептронов и меры ХИ-модуль // Информационно-управляющие системы. 2017. № 1 (86). С. 50–62. DOI:10.15217/issn1684-8853.2017.1.50.
14. Ложников П. С., Сулавко А. Е., Буряя Е. В., Писаренко В. Ю. Непрерывная биометрическая аутентификация пользователей в процессе работы на компьютере // Вопросы кибербезопасности. 2017. № 3. С. 24–34.
15. Min Rui, Choi Jongmoo, Medioni Gïrard [et al.]. Real-Time 3D Face Identification from a Depth Camera // Pattern Recognition (ICPR), 21st International Conference, 11–15 Nov 2012. P. 1739–1742.
16. Wang Xueqiao, Ruan Qiuqi, Jin Yi and An Gaojun. Three-dimensional face recognition under expression variation // EURASIP. Journal on Image and Video Processing. 2014. URL: <http://jivp.urasipjournals.com/content/2014/1/51> (дата обращения: 15.08.2017).
17. Кузнецов Л. А., Бугаков Д. А. Разработка меры оценки информационного расстояния между графическими объектами // Информационно-управляющие системы. 2013. № 1 (62). С. 74–79.
18. Hsu R. L., Abdel-Mottaleb M., Jain A. Face Detection In Color Images // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2002. Vol. 24. P. 696–706. DOI: 10.1109/34.1000242.
19. Martinkauppi B., Soriano M., Pietik'ainen M. Detection of Skin Color under Changing Illumination: A Comparative Study // Proceedings of the 12th International Conference on Image Analysis and Processing. 2003. P. 652–657.
20. Гермогенов А. П. Биометрические технологии идентификации личности и вопросы информационной безопасности их применения в государственных системах // Бизнес и безопасность в России. 2007. № 46. С. 45–49.
21. Zhao W., Chellappa R., Phillips P. [et al.]. Face recognition: a literature survey // ACM Computing Surveys. 2003. Vol. 35, Issue 4. P. 399–458.
22. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // Computer Vision and Pattern Recognition, 2001 (CVPR 2001). Proceedings of the 2001 IEEE. URL: [https://www.cs.utexas.edu/~grauman/courses/spring2007/395T/papers/viola\\_cvpr2001.pdf](https://www.cs.utexas.edu/~grauman/courses/spring2007/395T/papers/viola_cvpr2001.pdf) (дата обращения: 12.08.2017).
23. Cho H., Hwang SY. J. High-performance on-road vehicle detection with non-biased cascade classifier by weight-balanced training // EURASIP Journal on Image and Video Processing. 2015. Vol. 16. DOI:10.1186/s13640-015-0074-5.
24. Srinivasa K. G., Gosukonda S. Continuous multimodal user authentication: coupling hard and soft biometrics with support vector machines to attenuate noise // CSI Transactions on ICT, June 2014. Vol. 2, Issue 2. P. 129–140. DOI:10.1007/s40012-014-0054-4.
25. Kosilek R. P., Schopohl J., Grunke M. [et al.]. Automatic face classification of Cushing's syndrome in women – A novel screening approach // Experimental and Clinical Endocrinology and Diabetes. 2013. 121: P. 561–564. DOI:10.1055/s-0033-1349124.
26. Lawrence S., Giles C. L., Tsoi A. C. [et al.]. Face recognition: a convolutional neural-network approach // IEEE Transactions on neural networks. 1997. Vol. 8, № 1. P. 98–113.
27. Rybert Sipos I. Parallel stratified MCMC sampling of AR-HMMs for stochastic time series prediction / Ed. C. H. Skiadas // Proceedings, 4th Stochastic Modeling Techniques and Data Analysis International Conference with Demographics Workshop (SMTDA 2016). Valletta, 2016. P. 361–364.
28. Gorban A. N., Sumner N. R., Zinovyev A. Y. Beyond the concept of manifolds: principal trees, metro maps, and elastic cubic complexes // Eds. Gorban A. N. [et al.], Berlin, Heidelberg, Springer. LNCSE, 2007. Vol. 58. ISBN 978-3-540-73749-0. DOI: 10.1007/978-3-540-73750-6\_9.

ческого факультета Омского государственного технического университета (ОмГТУ).

Адрес для переписки: saret\_01@mail.ru

**ЕРЕМЕНКО Александр Валериевич**, кандидат технических наук, доцент кафедры «Инфокоммуникационные системы и информационная безопасность» Омского государственного университета путей сообщения. Адрес для переписки: nexus-@mail.ru

**СТАРКОВ Максим Андреевич**, студент гр. ИВТ-141 факультета «Информатика и вычислительная техника» ОмГТУ.

Адрес для переписки: gibsm319@gmail.com

**САМОТУГА Александр Евгеньевич**, научный сотрудник кафедры «Комплексная защита информации» радиотехнического факультета ОмГТУ.

Адрес для переписки: samotugasashok@mail.ru

Статья поступила в редакцию 19.09.2017 г.

© Г. А. Фофанов, А. В. Еременко, М. А. Старков, А. Е. Самотуга

## Книжная полка

**Никулин, Е. А. Компьютерная графика. Модели и алгоритмы : учеб. пособие / Е. А. Никулин. – СПб. : Лань, 2017. – 708 с. – ISBN 978-5-8114-2505-1.**

Рекомендовано Ученым советом Нижегородского государственного технического университета им. Р. Е. Алексеева в качестве учебного пособия для студентов направления подготовки «Информатика и вычислительная техника». В книге детально излагаются математические и алгоритмические основы современной компьютерной графики: модели графических объектов на плоскости и в пространстве (точки, векторы, линии и поверхности, включая составные, полиэдры, сплошные и воксельные объекты); геометрические задачи визуализации – комплекс алгоритмов 2d- и 3d-отсечения и удаления; алгоритмы аффинных и проективных преобразований; методы изображения поверхностей, включая текстурирование. Материал сопровождается большим числом иллюстраций, блок-схем алгоритмов и примеров их реализации. Настоящее пособие предназначено для студентов направления подготовки «Информатика и вычислительная техника». Также может быть полезно аспирантам, преподавателям вузов и всем специалистам, как постигающим основы компьютерной графики, так и разрабатывающим новые алгоритмы и прикладные графические программы.

**Фомичёв, В. М. Криптографические методы защиты информации. В 2 ч. Ч. 2. Системные и прикладные аспекты : учеб. / В. М. Фомичёв, Д. А. Мельников. – М. : Юрайт, 2017. – 246 с. – ISBN 978-5-534-01794-6, 978-5-534-01741-0.**

В учебнике систематически изложены разделы дискретной математики, используемые для изложения основ криптологии. Описаны способы построения современных криптографических систем, предназначенных для обеспечения информационной безопасности информационно-технологических систем, в том числе для защиты текстовых, речевых и факсимильных сообщений. Для закрепления знаний даны задачи, упражнения и контрольные вопросы. Издание состоит из двух частей. Первая часть посвящена математическим аспектам криптологии, вторая — системным и прикладным аспектам. Соответствует актуальным требованиям Федерального государственного образовательного стандарта высшего образования. Книга предназначена для студентов, аспирантов и преподавателей в области криптографических методов защиты информационного обмена в информационно-технических системах и сетях, а также для разработчиков и аналитиков систем криптографической защиты информации.

**Лаврищева, Е. М. Технология программирования и программная инженерия : учеб. / Е. М. Лаврищева. – М. : Юрайт, 2017. – 434 с. – ISBN 978-5-9916-8275-6.**

Серия «Университеты России» позволит высшим учебным заведениям нашей страны использовать в образовательном процессе учебники и учебные пособия по различным дисциплинам, подготовленные преподавателями лучших университетов России и впервые опубликованные в издательствах университетов. Все представленные в этой серии учебники прошли экспертную оценку учебно-методического отдела издательства и публикуются в оригинальной редакции. Описаны основные положения технологии программирования и инженерии программных продуктов и систем. Изложены отечественные и зарубежные методы разработки сложных программных систем из готовых компонентов повторного использования (КПИ). Рассмотрены подходы к инженерии программных продуктов из КИИ и их вариантов в SPLE (Software Product Line/Product Family), GDM (Generative Programming), Grid и др. Изложен метод сборки профамминных систем из КПИ с учетом модели характеристик (Feature Model), изменяющей структуру программных продуктов и систем. Описан объектно-компонентный метод моделирования вариантов программных систем и семейств программных систем из КПИ. Излагаются онтология представления доменов, методы извлечения знаний о готовых системах, спецификациях КПИ и доказательстве систем из них. Рассмотрены методы экспертизы, верификации, тестирования и оценки качества систем. Описаны дисциплины SE и методы создания веб-систем в среде Semantic Web и требования стандарта SEMAT к применению теории и методов SE в разработке систем. Для разработчиков и специалистов, которые интересуются современными технологиями изготовления отдельных программ и КИИ в разных языках программирования, их сборкой для получения вариантов программных систем и продуктов с обеспечением качества, а также студентов, магистров и специалистов по направлению подготовки «Теория и технология программирования».