

УДК 574:004.9

А.А. Фефелов, В.И. Литвиненко, М.А. Таиф, И.А. Лурье

## Гибридный подход при реконструкции генных регуляторных сетей

Описан гибридный метод реконструкции генной регуляторной сети по временным рядам данных экспрессии генов. В качестве модели сети предложена система обыкновенных дифференциальных уравнений. Для проверки эффективности метода проведены исследования на двух моделях искусственных регуляторных сетей.

**Ключевые слова:** генные регуляторные сети, реверсная инженерия, экспрессия генов, обыкновенные дифференциальные уравнения, алгоритм клонального отбора, вейвлет-нейронная сеть.

Розроблено гібридний метод реконструкції генної регуляторної мережі у часових рядах даних експресії генів. Як модель запропоновано систему звичайних диференціальних рівнянь. Для перевірки ефективності методу проведено дослідження на двох моделях штучних регуляторних мереж.

**Ключові слова:** генні регуляторні мережі, реверсна інженерія, експресія генів, звичайні диференціальні рівняння, алгоритм клонального відбору, вейвлет-нейронна мережа.

**Введение.** Реверсная инженерия *генных регуляторных сетей* (ГРС) (реконструкция) представляет собой процесс восстановления структурных и динамических характеристик сети на основе данных наблюдения. Информация о регуляторных контурах генов, получаемая в результате реконструкции ГРС, находит широкое применение в медицине и биологии, когда можно обнаружить генетические заболевания, разработку новых лекарственных препаратов, снижение побочных эффектов от лечения, формирование гипотез о функциональном назначении неизвестных генов [1, 2] и пр. Наиболее распространенным источником данных для реконструкции ГРС служат профили экспрессии генов, формируемые посредством ДНК-микрочипов, что позволяет одновременно измерять уровни экспрессии множества генов, создавая внушительные объемы данных, ранее недоступных для исследования [3].

### Методология реконструкции генных регуляторных сетей

Данные временного ряда ДНК-микрочипов представляют собой числовую матрицу, состоящую из тысяч строк (указывающих на гены) и десятков столбцов (временных точек). Общий вид этих данных выглядит так [4]:

$$\begin{array}{cccc} & t_1 & t_2 & \dots & t_T, \\ g_1 & D_{11} & D_{12} & \dots & D_{1T}, \\ g_2 & D_{21} & D_{22} & \dots & D_{2T}, \\ \vdots & \vdots & \vdots & & \\ g_n & D_{n1} & D_{n2} & \dots & D_{nT}, \end{array}$$

где  $g_i$  – гены, а  $t_j$  – моменты времени.  $D_{ij}$  – значения экспрессии  $i$ -го гена в  $j$ -й момент времени. Таким образом, значение сигнала  $g_i(t)$  в момент времени  $t = j$  равно  $D_{ij}$ . В генных сетях значение  $g_i(t)$  связано с самим собой и другими генами при  $t-1$ . Следовательно, целью методов моделирования генных регуляторных сетей является изучение связей между генами, определяющими динамику и структуру сети.

Существует множество различных моделей и методов реконструкции ГРС, имеющих достоинства и недостатки. Среди них можно выделить модели булевых сетей [5], в которых уровень экспрессии генов представлен бинарными значениями (ноль и единица), а отношения между генами заданы в форме булевых операций; модели байесовских сетей [6], где связи между генами рассматриваются как статистические зависимости и модели, описывае-

мые обыкновенными дифференциальными уравнениями (ОДУ) которые показывают изменения концентрации друг с другом и с внешним возмущением [7]. Под внешним возмущением понимаем экспериментальную обработку, которая может изменить скорость транскрипции генов. Примером возмущения может быть воздействие химического соединения, например лекарства или генетическое возмущение, которое вызывает избыточную экспрессию или репрессию определенных генов. Упомянутые уравнения – эффективный и гибкий инструмент, способный к выявлению сложных динамических взаимосвязей между компонентами ГРС. Однако актуальной остается проблема получения адекватной модели, что представляется достаточно сложной задачей. В связи с этим во многих исследованиях используется фиксированная запись уравнений, полученная либо эмпирическим путем, либо на основании фундаментальных исследований и экспертных знаний. При этом задача реконструкции сводится к параметрической оптимизации выбранных выражений. Одним из наиболее популярных примеров такого подхода является S-система [8]. Существенным недостатком фиксированной записи уравнений можно считать отсутствие гибкости модели, что ограничивает область ее применения.

Цель данной статьи – показать разработку гибридной процедуры решения проблемы реконструкции генных сетей на основе ОДУ. Для устранения упомянутого недостатка использована общая форма системы дифференциальных уравнений, которую можно представить так:

$$\begin{aligned} \frac{dx_i}{dt} &= f_i(x_1, x_2, \dots, x_n, u, \theta_i), \\ f_i(0, \dots, 0) &= 0, \quad i = 1, 2, \dots, n, \end{aligned} \quad (1)$$

где  $\theta_i$  – набор параметров, описывающих взаимодействия между генами,  $x_i(t)$  – переменная концентрации продукта экспрессии  $i$ -го гена в момент времени  $t$ ;  $\dot{x}_i(t) = \frac{dx_i}{dt}(t)$  – скорость изменения концентрации  $i$ -го гена,  $n$  –

количество генов в сети,  $u$  – внешнее возмущение системы.

Поскольку ОДУ детерминированы, взаимодействия между генами ( $\theta_i$ ) представляют собой причинные взаимодействия, а не статистические зависимости, как в других методах.

Для реконструкции сети с использованием ОДУ, необходимо выбрать функциональную форму для  $f_i$  и оценить неизвестные параметры  $\theta_i$  для каждого  $i$ -го гена с использованием некоторой методики оптимизации. Подходы, основанные на ОДУ, позволяют получать ориентированные графы и могут применяться как для стационарных экспрессий генов, так и экспрессий генов, описываемых многомерными временными рядами. Другим преимуществом ОДУ является возможность использования уравнения (1) для прогнозирования поведения сети в различных условиях, когда параметры  $\theta_i$  для всех  $i$ -х генов известны, т.е. удаление гена, лечение посредством какого-либо внешнего химического или физического агента.

Для структурной идентификации правой части системы ОДУ предложен гибридный вычислительный метод, основанный на комбинировании искусственной иммунной системы и нейронной сети.

Идея объединения различных вычислительных методов и образования гибридов основана на предположении о том, что полученный в результате объединения новый вычислительный метод должен обладать более высокой производительностью, чем его составляющие. Эта идея привела к созданию технологии гибридизации, позволяющей синтезировать целые классы алгоритмов, способных к решению сложных задач на качественно новом уровне. Согласно данной технологии, сильные стороны методов, использованных в гибридизации, формируют комбинированный результат, характеризующий основные преимущества гибридного подхода, такие как получение более качественных решений; получение решений за меньшее время; решение задач большой размерности. В нашем случае технология гибридизации предполагает использование вычис-

лительной модели вейвлет-нейронной сети, в качестве метода обучения которой выбран один из видов искусственных иммунных систем, – алгоритм клонального отбора.

### Постановка задачи

Для временного ряда  $X(t)$  данных экспрессии генов в генной регуляторной сети, построить модель, представленную в виде системы ОДУ (1), такую чтобы среднеквадратическая ошибка модели на данных  $X(t)$  была минимальной. Задача предусматривает выбор оптимальной структуры и параметров модели.

Концептуальная схема предложенной технологии реконструкции ГРС показана на рис. 1.

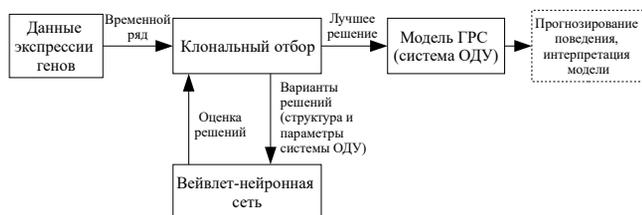


Рис. 1

Эффективность данной технологии покажем путем проведения экспериментов.

### Алгоритм клонального отбора

Современные исследования в области иммунологии позволили сформировать особый взгляд на естественную иммунную систему животных и человека и рассматривать ее как мощную вычислительную структуру. Дальнейшее развитие иммунной вычислительной парадигмы привело к созданию целого ряда методов под общим названием – *искусственные иммунные системы* (ИИС) [9, 10]. Специфические свойства этой системы как распределенной, децентрализованной и неоднородной структуры положены в основу различных моделей, описывающих процессы поиска, обнаружения, распознавания и защиты организма от чужеродных агентов – вирусов и бактерий. Одна из таких моделей – клональный отбор [11, 12], функция которой заключается в последовательной адаптации популяции антител – вариантов решения задачи к входящему антигену – целевой функции.

В общем виде, формально алгоритм клонального отбора можно представить так [13]:

$$CLONALG = (Ab^0, f, L, N, n, \partial, d, e), \quad (2)$$

где  $Ab^0$  – исходная популяция антител;  $f$  – функция аффинности, которая строится на основе целевой функции ( $f \Rightarrow \max$  или  $f \Rightarrow \min$ );  $N$  – количество антител в популяции;  $L$  – длина рецептора антитела;  $n$  – количество антител, отбираемых для клонирования (с самой высокой аффинностью);  $\partial$  – множительный фактор, регулирующий количество клонов отобранных антител;  $d$  – количество антител, подлежащих замене новыми (с самой низкой аффинностью);  $e$  – критерий останова.

Под аффинностью связей антител между собой или антител с антигенами понимается расстояние между соответствующими векторами атрибутов, которое выражается в виде скалярного неотрицательного значения:  $P^l \times P^l \rightarrow \mathcal{R}^+$ , определяющего степень соответствия между молекулами (векторами атрибутов). Причем при использовании оговоренной ранее степени подобия форм получается, что чем меньше расстояние между индивидуумами, тем выше их аффинность друг к другу. Цель решения подобного рода задач состоит в поиске оптимальных значений (минимумов или максимумов) некоторого критерия  $y = f(x_1, x_2, \dots, x_l)$ ,  $x_i \in X$ ,  $i = \overline{1, l}$ , где  $X$  – допустимое множество задачи. В общем случае рассматриваются задачи многокритериальной оптимизации:

$$y = (y_1, y_2, \dots, y_n) \rightarrow \min (\max), \quad (3)$$

где  $y_j = f_j(x_1, x_2, \dots, x_l)$ ,  $j = \overline{1, n}$ ,  $n$  – количество критериев задачи. В зависимости от условий задачи возможен поиск глобального или локальных оптимумов.

В задачах оптимизации обобщенная форма антител представляет собой вектор аргументов  $Ab = (x_1, x_2, \dots, x_l)$ , а в качестве антигенов используются собственно критерии  $y_j$ , выраженные в виде функций:  $Ag = f(x_1, x_2, \dots, x_l)$ . Значения аффинности  $g_j$  вычисляются на основе значений критериев  $y_j$ , отображенных во множество неотрицательных чисел, т.е.:

$$f: X \rightarrow \mathcal{R}, \quad F: \mathcal{R} \rightarrow \mathcal{R}^+. \quad (4)$$

Таким образом, имеется некоторая функция аффинности  $g = F(f(x_1, x_2, \dots, x_n))$ , определяющая степень соответствия индивидуумов друг другу. В таких задачах нельзя оперировать понятием расстояния, так как оптимальные значения критериев заранее не известны, и, следовательно, не известна максимально возможная степень соответствия индивидуумов. Поэтому управление динамикой ИИС осуществляется с учетом относительных значений аффинности или ранга индивидуумов совокупности.

Псевдокод алгоритма клонального отбора показан на рис. 2.

```

Поколение = 0
Создать начальную популяцию антител ( $Ab^0$ )
цикл пока условие_останова  $e = \text{false}$ 
    поколение = поколение + 1
    Вычислить аффинность антител к антигену ( $f$ )
    Выбрать антитела с наибольшей аффинностью
    Создать клоны выбранных антител (в количестве  $n-f$ )
    Провести мутацию клонов (с вероятностью  $p_{mut}^{1-f}$ )
    Вычислить аффинность клонов к антигену ( $f_2$ )
    Выбрать клоны с наибольшей аффинностью
    Перенести клоны в основную популяцию
    Заменить  $d$  антител с наименьшей аффинностью
конец цикла пока

```

Рис. 2

Несмотря на существование множества модификаций алгоритма, все они имеют сходный принцип работы: популяция антител подвергается некоторым преобразованиям, в результате которых антитела повышают свою аффинность.

В ИИС в зависимости от типа задачи можно использовать различные способы представления решений. Наиболее часто используется бинарное представление, при котором антитело формируется как непрерывная строка бит, состоящая из значений аргументов (независимых переменных) или параметров задачи. Условия и цели задачи также являются определяющими при выборе способа представления иммунных операторов, вида функции аффинности, значений параметров алгоритма.

При вычислении аффинности основной популяции создаются условия для отбора тех клеток, которые максимально полно (на дан-

ном этапе) входят во взаимодействие с антигеном. В процессе активации, отобранные антитела увеличивают свое представительство в пространстве решений в результате клонирования. Клетки, аффинность которых выше, создают большее количество клонов, но меньше подвержены мутации. Мутация в ИИС имеет высокую интенсивность, так как служит главной движущей силой эволюции. В процессе замены клетки с низкой аффинностью удаляются из основной популяции, а на их место приходят новые, случайно сгенерированные индивидуумы. Это позволяет исследовать всю целевую поверхность, а не оставаться в локальных экстремумах.

### Вейвлет-нейронные сети

В настоящее время среди множества приложений, использующих вейвлет-функции, повышенный интерес вызывают вейвлет-нейронные сети, разработанные для решения задач нелинейного регрессионного анализа [14–16]. Вейвлет-нейронные сети можно применять как для построения статических моделей, так и для моделирования динамики нелинейных нестационарных систем. Доказано, что семейства вейвлетов являются универсальными аппроксиматорами и поэтому могут использоваться при моделировании динамических процессов. Данное свойство говорит о том, что любая функция может быть аппроксимирована с любой заданной точностью при помощи конечной суммы вейвлетов. Таким образом, вейвлет-нейронные сети выступают как альтернатива перцептронам и радиально-базисным нейронным сетям.

При решении задач синтеза и обучения вейвлет-нейронных сетей основное внимание уделяется настройке параметров вейвлет-нейронов, расположенных в скрытом слое сети, что, по сути, является задачей оптимизации. В данной статье предложено использовать алгоритм клонального отбора для оптимизации параметров вейвлет-нейронной сети в процессе ее обучения [17].

Архитектура вейвлет-нейронной сети представлена на рис. 3.

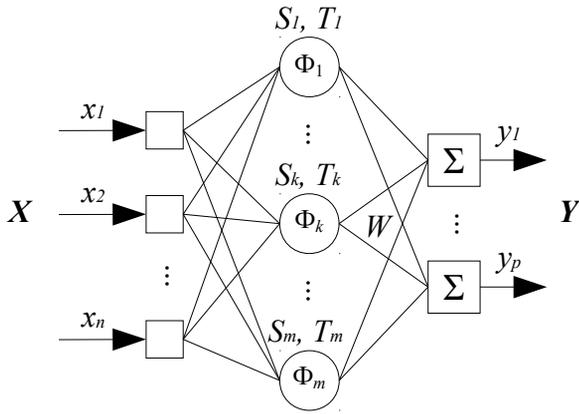


Рис. 3

Нейроны скрытого слоя – это вейвлет-нейроны ( $\Phi$ ), содержащие настраиваемый вейвлет. В качестве параметров вейвлета используется его масштаб ( $s$ ) и сдвиг ( $t$ ) по временной оси. Вейвлет-нейронные сети используют семейства вейвлетов, образованных от одного материнского посредством применения к нему операций масштабирования и сдвига. Например, материнский вейвлет одного из наиболее популярных типов, а именно мексиканской шляпы, выглядит так:

$$\phi(z) = (z^2 - 1)e^{-z^2/2}. \quad (5)$$

На основе данного материнского вейвлета можно построить семейство производных при помощи выражения:

$$\phi'(z) = \phi\left(\frac{z-t}{s}\right). \quad (6)$$

Построение оптимального семейства производных вейвлетов является основной задачей обучения вейвлет-нейронной сети.

Основываясь на представленной архитектуре, выход нейронной сети будет определяться следующим образом:

$$y_j = \sum_{k=1}^m w_{kj} \Phi_k(X, S_k, T_k), \quad j = 1, 2, \dots, p, \quad (7)$$

где  $X = (x_1, x_2, \dots, x_n)$  – произвольный  $n$ -мерный вектор входных переменных;  $w_{kj} \in W$  – веса линейного слоя;  $S_k = (s_{1k}, s_{2k}, \dots, s_{nk})$  и  $T_k = (t_{1k}, t_{2k}, \dots, t_{nk})$  соответственно векторы масштаба и сдвига  $k$ -го вейвлета скрытого слоя сети относительно вектора входных переменных;  $p$  – количество выходов нейронной сети.

Поскольку моделируемый процесс содержит множество входных переменных, то вейвлеты  $\Phi_k$  – многомерны. Для построения многомерного вейвлета использована следующая формула [15]:

$$\Phi_k(X, S_k, T_k) = \prod_{i=1}^n \phi_k\left(\frac{x_i - t_{ik}}{s_{ik}}\right), \quad (8)$$

где  $s_{ik} \in S_k, t_{ik} \in T_k$  – элементы векторов масштаба и сдвига для каждой входной переменной сети.

### Метод реконструкции системы ОДУ

В данном случае правая часть системы ОДУ представлена нейронной сетью. Это означает, что на вход сети поступают временные ряды изменения концентрации продуктов экспрессии генов, а выход сети соответствует вектору первых производных концентраций тех же продуктов (рис. 4). Очевидно, что число выходов такой нейронной сети будет всегда равно числу ее входов, т.е.  $p = n$ .

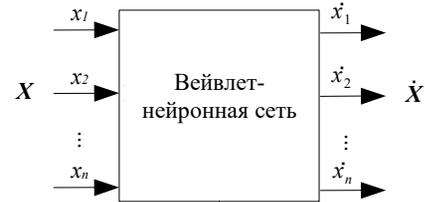


Рис. 4

Построив данную сеть, можно будет найти форму уравнений, которая лучше всего соответствует данным экспрессии. Синтез и обучение вейвлет-нейронной сети выполняется при помощи алгоритма клональной селекции. Для оптимизации выбраны следующие параметры сети: количество нейронов скрытого слоя ( $m$ ), типы вейвлетов скрытого слоя ( $\phi$ ), значения сдвига для нейронов скрытого слоя ( $t_{ik}$ ), значения масштаба для нейронов скрытого слоя ( $s_{ik}$ ), веса выходного слоя ( $w_{kj}$ ), параметры функции и активации выходного слоя ( $a_j$ ). Исходя из перечня выделенных параметров, получаем следующую структуру индивидуума антитела ИИС, кодирующего вейвлет-нейронную сеть (рис. 5).

Для кодирования значений параметров использована двоичная система. Точность представления (количество бит на одно значение)

$b_1$	...	$b_m$	$\phi_1$	...	$\phi_m$	$t_{11}$	...	$t_{1n}$	$s_{11}$	...	$s_{1n}$	...	$t_{m1}$	...	$t_{mn}$	$s_{m1}$	...	$s_{mn}$	$w_{11}$	...	$w_{m1}$	...	$w_{1n}$	...	$w_{mn}$	$a_1$	...	$a_n$
-------	-----	-------	----------	-----	----------	----------	-----	----------	----------	-----	----------	-----	----------	-----	----------	----------	-----	----------	----------	-----	----------	-----	----------	-----	----------	-------	-----	-------

Рис. 5

выделена в качестве параметра настройки ИИС. Элементы строки  $b_1, \dots, b_m$  кодируют состояния соответствующих вейвлет-нейронов скрытого слоя. Значение *ноль* соответствует выключенному состоянию (нейрон не участвует в расчете выходного значения нейронной сети). Значение *единица* – нейрон включен. Благодаря такой схеме обеспечивается автоматический поиск оптимального количества скрытых элементов нейронной сети. В настройках ИИС устанавливается лишь максимально возможное количество этих элементов.

Элементы  $\phi_1, \dots, \phi_m$  соответствуют типу вейвлета, заданного индексом в массиве материнских вейвлетов. В данной статье использованы вейвлеты из следующих семейств: *mexican hat, morlet, polywog, rasp, slog, shannon*. Обучающий алгоритм может выбирать любой из вейвлетов в качестве нейрона скрытого слоя, вследствие чего в рамках одной нейронной сети могут быть использованы представители различных типов.

Для расчета аффинности индивидуумов выбрано следующее выражение:

$$f = \text{MIN} \sqrt{\frac{1}{n \cdot T} \sum_{i=1}^n \sum_{j=0}^{T-1} (x_i^M(t_0 + j\Delta t) - x_i(t_0 + j\Delta t))^2}, \quad (9)$$

где  $t_0$  – начальное время;  $\Delta t$  – временной шаг;  $n$  – количество компонентов (генов) в ГРС;  $T$  – количество данных временного ряда экспрессии;  $x_i^M(t_0 + j\Delta t)$  – значения концентрации, полученные решением системы ОДУ, представленной вейвлет-нейронной сетью;  $x_i(t_0 + j\Delta t)$  – значения концентрации из временного ряда экспрессии.

По сути (9) – это среднеквадратическая ошибка модели, вычисленная в одном выражении для временных рядов экспрессии всех компонентов регуляторной сети.

Значения  $x_i^M$  для каждого индивидуума популяции ИИС получаются решением системы ОДУ одним из численных методов с начальными условиями  $x_i(t_0)$ . В статье в качестве чис-

ленного метода решения задачи Коши выбран метод Рунге–Кутты четвертого порядка.

Далее приведено обобщенное пошаговое описание предложенного алгоритма.

Шаг 1. *Инициализация*. Создание случайной начальной популяции антител ( $Ab^0$ ).

Шаг 2. *Вычисление аффинности*. Для каждого антитела  $Ab_i$ ,  $Ab_i \in Ab$  вычислить аффинность согласно выражению (9) с использованием метода Рунге–Кутты.

Шаг 3. *Клональная селекция и распространение*. Выбрать из популяции  $n$  лучших антител и поместить их в отдельную популяцию клонов  $Ab_C$ . Генерировать клоны элементов популяции  $Ab_C$  пропорционально их аффинности, т.е. чем выше аффинность, тем большее количество клонов создается, и наоборот. В статье реализована селекция по принципу турнира [17], что дает возможность управлять сходимостью алгоритма и поддерживать разнообразие популяции на необходимом уровне.

Шаг 4. *Созревание аффинности*. Подвергнуть мутации все клоны популяции  $Ab_C$  с вероятностью обратно пропорциональной их аффинностям, т.е. чем ниже аффинность индивидуума, тем выше вероятность его мутации. Вычислить новую аффинность каждого антитела  $Ab_j$ ,  $Ab_j \in Ab_C$  аналогично шагу 2. Выбрать из популяции  $Ab_C$  (при помощи турнира)  $n$  лучших антител и перенести их в популяцию клеток памяти  $M_R$ . В статье применена побитовая мутация индивидуумов, согласно которой каждый бит антитела может быть изменен с заданной вероятностью.

Шаг 5. *Метадинамика*. Заменить  $d$  худших антител популяции  $Ab$  новыми случайными индивидуумами.

Шаг 6. Заменить  $n$  антител популяции  $Ab$  клетками памяти из  $M_R$  и перейти к шагу 2 пока не будет достигнут критерий останова. В качестве критерия останова выбрано достижение популяцией определенного количества поколений.

## Эксперименты

Для экспериментов выбраны две задачи разной степени сложности [18]. Для каждой из них сгенерированы по три набора данных временно-го ряда с различными начальными условиями.

В первом эксперименте используется простая система ОДУ следующего вида:

$$\begin{cases} \dot{x}_1 = -0,9x_2 + x_3 \\ \dot{x}_2 = 0,2x_1x_2 \\ \dot{x}_3 = x_3 - 0,5x_1 \end{cases} \quad (10)$$

Для каждого из трех временных рядов первого эксперимента выбраны следующие начальные условия:  $(x_1, (t_0), x_2, (t_0), x_3, (t_0)) = (0,2; 0,1; 0,3), (0,1; 0,1; 0,1), (0,2; 0,2; 0,2)$ .

Параметры алгоритма реконструкции, используемые в данном эксперименте, представлены в табл. 1.

Таблица 1

Название параметра	Значение параметра
Размер основной популяции	100
Размер популяции клонов	600
Количество поколений	900
Коэффициент отбора	0,7 ( $n=70$ )
Размер турнира	5
Уровень мутации	0,01
Количество нейронов	7
Точность представления	8 бит
Временной шаг	0,25
Количество данных ряда	20

Среднеквадратические ошибки обучения на каждом наборе данных составили соответственно:  $f_1 = 0,0216, f_2 = 0,0344, f_3 = 0,02$ . График временных рядов лучшей модели, синтезированной в результате эксперимента с начальными условиями  $(0,2; 0,1; 0,3)$ , показан на рис. 6.

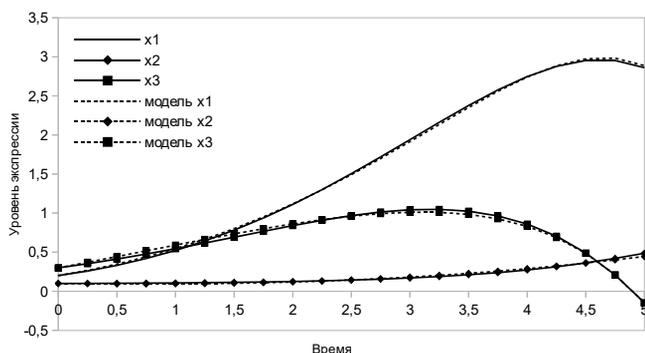


Рис. 6

Модель, соответствующая рис. 6, представлена в (11).

$$\begin{cases} \dot{x}_1 = 5,665\Phi_1 - 5,994\Phi_2 - 0,776\Phi_3 - \\ \quad - 0,353\Phi_4 - 3,785\Phi_5 - 5,994\Phi_6 \\ \dot{x}_2 = 0,384\Phi_1 - 0,008\Phi_2 + 0,557\Phi_3 + \\ \quad + 0,369\Phi_4 - 2\Phi_5 - 1,529\Phi_6 \\ \dot{x}_3 = 1,336\Phi_1 - 4,746\Phi_2 - 4,993\Phi_3 - \\ \quad - 4,541\Phi_4 + 5,239\Phi_5 - 3,596\Phi_6 \end{cases} \quad (11)$$

где:

$$\Phi_1 = SLOG2\left(\frac{x_1 - 0,51}{5}\right) \cdot SLOG2\left(\frac{x_2 + 0,071}{5,176}\right) \times \\ \times SLOG2\left(\frac{x_3 - 1,388}{24,41}\right) \quad (12)$$

$$\Phi_2 = POLYWOG2\left(\frac{x_1 + 1,78}{7,647}\right) \cdot POLYWOG2\left(\frac{x_2 + 1,937}{7,294}\right) \times \\ \times POLYWOG2\left(\frac{x_3 - 0,792}{5,706}\right) \quad (13)$$

$$\Phi_3 = POLYWOG1\left(\frac{x_1 + 0,337}{10,65}\right) \cdot POLYWOG1\left(\frac{x_2 - 2}{5}\right) \times \\ \times POLYWOG1\left(\frac{x_3 - 1,953}{5,353}\right) \quad (14)$$

$$\Phi_4 = POLYWOG2\left(\frac{x_1 - 1,357}{11,71}\right) \cdot POLYWOG2\left(\frac{x_2 - 2}{7,824}\right) \times \\ \times POLYWOG2\left(\frac{x_3 + 1,686}{5,71}\right) \quad (15)$$

$$\Phi_5 = POLYWOG1\left(\frac{x_1 - 1,075}{5}\right) \cdot POLYWOG1\left(\frac{x_2 + 1,89}{5,529}\right) \times \\ \times POLYWOG1\left(\frac{x_3 - 1,984}{5,353}\right) \quad (16)$$

$$\Phi_6 = RASP\left(\frac{x_1 - 0,0235}{35,88}\right) \cdot RASP\left(\frac{x_2 + 0,353}{33,41}\right) \times \\ \times RASP\left(\frac{x_3 + 1,325}{12,24}\right) \quad (17)$$

В данной модели функции *SLOG2*, *POLYWOG1*, *POLYWOG2*, *RASP1* – материнские вейвлеты [19].

В основу второго эксперимента положена ГРС, полученная искусственным путем и представленная в виде направленного взвешенного графа (рис. 7) [20]. В этой сети узлы графа являются генами, а дуги показывают направления регуляторных воздействий и их уровень. Положительный вес дуги говорит об активирующем воздействии, отрицательный – о подавляющем воздействии.

На основе данной сети, согласно [20], построены временные ряды экспрессии, в которых уровень экспрессии  $i$ -го гена  $x_i(t+1)$  вычисляется так:

$$x_i(t+1) = \frac{m_i}{1 + e^{-s_i(t)}}, \quad (18)$$

$$s_i(t) = \sum_{j=0}^n w_{ij} x_j(t), \quad (19)$$

где  $m_i$  – максимальный уровень экспрессии  $i$ -го гена (для нашего эксперимента принято  $m_1 = 4,0$ ;  $m_2 = 5,0$ ;  $m_3 = 5,0$ );  $n$  – количество генов в регуляторной сети;  $w_{ij}$  – уровень влияния гена  $j$  на ген  $i$ . Уровень влияния задается матрицей связности (табл. 2).

Таблица 2

	$x_1$	$x_2$	$x_3$
$x_1$	0,9	0	-0,8
$x_2$	0,3	0	0
$x_3$	0	0,9	0

Для каждого из трех временных рядов второго эксперимента выбраны следующие начальные условия:  $(x_1(t_0), x_2(t_0), x_3(t_0)) = (0,5; 0,0; 0,0)$ ,  $(0,0; 0,5; 0,0)$ ,  $(0,0; 0,0; 0,5)$ .

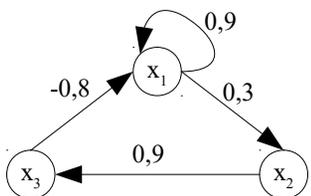


Рис. 7

Параметры алгоритма реконструкции, используемые в данном эксперименте, представлены в табл. 3.

Таблица 3

Параметры	Значения параметров
Размер основной популяции	300
Размер популяции клонов	900
Количество поколений	900
Коэффициент отбора	0,7 ( $n=210$ )
Размер турнира	8
Уровень мутации	0,01
Количество нейронов	7
Точность представления	8 бит
Временной шаг	0,25
Количество данных ряда	36

Среднеквадратические ошибки обучения в каждом наборе данных составили соответственно:  $f_1 = 0,059$ ;  $f_2 = 0,05$ ;  $f_3 = 0,063$ . График временных рядов лучшей модели, синтезированной во втором эксперименте с начальными условиями  $(0,5; 0,0; 0,0)$ , показан на рис. 8.

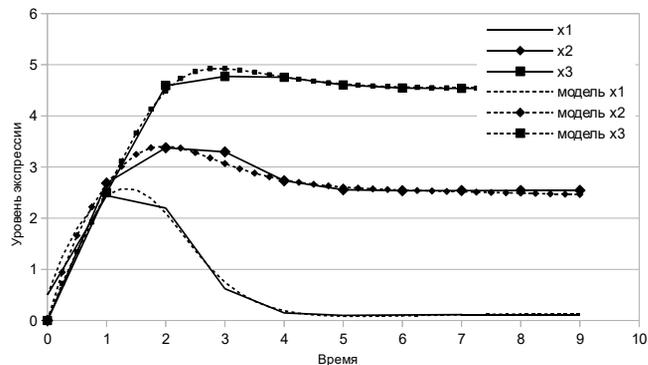


Рис. 8

Представим модель, соответствующую рис. 8.

$$\begin{cases} \dot{x}_1 = 11,121\Phi_1 - 9,88\Phi_2 - 6,466\Phi_3 - \\ \quad - 6,673\Phi_4 - 13,19\Phi_5 + 6,983\Phi_6 \\ \dot{x}_2 = 1,883\Phi_1 - 6,924\Phi_2 - 7,098\Phi_3 + \\ \quad + 3,679\Phi_4 + 0,029\Phi_5 + 4,259\Phi_6 \\ \dot{x}_3 = -5,657\Phi_1 - 12,265\Phi_2 - 6,758\Phi_3 + \\ \quad + 10,463\Phi_4 - 9,662\Phi_5 + 9,562\Phi_6 \end{cases}, \quad (20)$$

где:

$$\Phi_1 = \text{POLYWOG2}\left(\frac{x_1 + 0,059}{3,976}\right) \cdot \text{POLYWOG2}\left(\frac{x_2 + 4,725}{50,52}\right) \times \text{POLYWOG2}\left(\frac{x_3 - 0,333}{21,29}\right), \quad (21)$$

$$\Phi_2 = \text{POLYWOG2}\left(\frac{x_1 - 5}{35,64}\right) \cdot \text{POLYWOG2}\left(\frac{x_2 - 2,529}{3,165}\right) \times \text{POLYWOG2}\left(\frac{x_3 - 4,529}{18,32}\right), \quad (22)$$

$$\Phi_3 = \text{SHANNON}\left(\frac{x_1 - 2,608}{68,92}\right) \cdot \text{SHANNON}\left(\frac{x_2 - 5}{70}\right) \times \text{SHANNON}\left(\frac{x_3 + 3,941}{43,75}\right), \quad (23)$$

$$\Phi_4 = \text{POLYWOG1}\left(\frac{x_1 - 1,118}{18,32}\right) \cdot \text{POLYWOG1}\left(\frac{x_2 + 0,294}{4,788}\right) \times \text{POLYWOG1}\left(\frac{x_3 + 4,255}{12,36}\right), \quad (24)$$

$$\Phi_3 = \text{POLYWOG5}\left(\frac{x_1 - 4,176}{36,18}\right) \cdot \text{POLYWOG5}\left(\frac{x_2 + 4,059}{18,32}\right) \times \text{POLYWOG5}\left(\frac{x_3 + 3,667}{35,64}\right), \quad (25)$$

$$\Phi_6 = \text{SHANNON}\left(\frac{x_1 + 0,02}{70}\right) \cdot \text{SHANNON}\left(\frac{x_2 + 0,02}{5,6}\right) \times \text{SHANNON}\left(\frac{x_3 + 5}{28,06}\right). \quad (26)$$

В данной модели функции *SHANNON* и *POLYWOG5* – материнские вейвлеты [19].

**Заключение.** Предложенный гибридный метод реконструкции правой части системы обыкновенных дифференциальных уравнений используется для описания динамики генных регуляторных сетей. В качестве вычислительной модели выбрана вейвлет-нейронная сеть, обучаемая посредством алгоритма клональной селекции. В качестве обучающих данных применяются профили экспрессии генов, представленные временными рядами изменения концентрации продуктов экспрессии.

Эффективность предложенного метода показана экспериментальными исследованиями, подтверждающими применимость данного подхода к поиску взаимосвязей между компонентами ГРС. Несмотря на использование сетей, состоящих из небольшого количества генов, можно рассчитывать на то, что дальнейшее совершенствование технологии в совокупности с методами предобработки позволит проводить эффективную реконструкцию реальных ГРС. Основными направлениями дальнейших исследований выбрано создание метапроцедур автоматической настройки параметров гибридного алгоритма, что позволит сократить пространство поиска благодаря динамическому изменению интервалов представления значений элементов, составляющих индивидуумы ИИС.

1. Rawool S.B., Venkatesh K.V. Steady state approach to model gene regulatory networks – simulation of microarray experiments // *Biosystems*, 2007, **90**(3), P. 636–655.
2. Chen P.C.Y., Chen J.W. A Markovian approach to the control of genetic regulatory networks // *Ibid*, **90**(2), P. 535–545.
3. DeRisi J.L., Lyer V.R., Brown P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale // *Science*, 1997, **278**, P. 680–686.

4. A review of modeling techniques for genetic regulatory networks / H. Yaghoobi, S. Haghypour, H. Hamzeiy et al. // *J. Med. Sign. Sens*, 2012, **2** (1), P. 61–70.
5. Liang S., Fuhrman S., Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures // *Pacific Symp. on Biocomputing (PSB '98)*, Jan. 1998: Proc., Maui, Hawaii, USA, 1998, P. 18–29.
6. Using Bayesian networks to analyze expression data / N. Friedman, M. Linial, I. Nachman et al. // *J. of Computational Biology*, 2000, **7**(3–4), P. 601–620.
7. Chen T., He H.L., Church G.M. Modeling gene expression with differential equations // *4th Pacific Symp. on Biocomputing (PSB '99)*, Jan. 1999: Proc., Big Island of Hawaii, USA, 1999, P. 29–40.
8. Savageau M.A. Introduction to S-systems and the underlying power-law formalism // *Mathematical and Computer Modelling*, 1988, **11**, P. 546–551.
9. Dasgupta D. *Artificial Immune Systems and Their Applications*, Springer-Verlag, 1999, 306 p.
10. De Castro L.N., Timmis J. *Artificial Immune Systems: A New Computational Intelligence Approach*, Springer, Heidelberg, 2002, 357 p.
11. De Castro L.N., Von Zuben F.J. *Artificial Immune Systems: Part I – Basic Theory and Applications*, (Technical Report – RT DCA 01/99, FEEC/UNICAMP), Brazil, 1999, 95 p.
12. De Castro L.N., Von Zuben F.J. *Artificial Immune Systems: Part II – A Survey of Applications*, (Technical Report – RT DCA 02/00, FEEC/UNICAMP), Brazil, 2000, 64 p.
13. Бідюк П.І., Литвиненко В.І., Фефелов А.О. Формалізація методів побудови штучних імунних систем // *Наук. вісті Нац. техн. ун-ту України «КПІ»*, 2007, № 1, С. 29–41.
14. Linear-Wavelet Networks / R.K.H Galvao, V.M. Baccera, J.M.F. Calado et al. // *J. of Appl. Math. Comput. Sci.*, 2004, **14**, № 2, P. 221–232.
15. Oussar Y., Dreyfus G. Initialization by Selection for Wavelet Network Training // *Neurocomputing*, 2000,–**34**, P. 131–143.
16. Training Wavelet Networks for Nonlinear Dynamic Input-Output Modeling / Y. Oussar, I. Rivals, L. Personnaz et al. // *Ibid*, 1998, **20**, P. 173–188.
17. Фефелов А.А., Горавский С.П., Литвиненко В.И. Построение и настройка вейвлет-нейронных сетей при помощи искусственных иммунных систем // *Зб. наук. праць: У 5 т. Міжнар. наук. конф. «Інтелектуальні системи прийняття рішень та прикладні аспекти інформаційних технологій»*, Євпаторія, 2005, Т. 5, С. 77–80.
18. Goldberg D.E., Deb K. A comparative analysis of selection schemes used in genetic algorithms // *Foundations of genetic algorithms*, Morgan Kaufmann, 2014, 341 p.

19. *Sakamoto E., Iba H.* Inferring a system of differential equations for a gene regulatory network by using genetic programming // Congress on Evolutionary Computation, 27–30 May 2001: Proc., COEX, Seoul, Korea, 2001, P. 720–726.
20. *Lekutai G.* Adaptive self-tuning neuro wavelet network controllers: thesis... doctor of philosophy in «The Electrical Engineering Department», Blacksburg, Virginia, 1997, 112 p.
21. *Weaver D.C., Workman C.T., Stormo G.D.* Modeling regulatory networks with weight matrices // Pacific Symp. on Biocomputing 4 (PSB '99), 4–9 Jan. 1999: Proc., Maui, Hawaii, USA, 1999, P. 112–123.

Поступила 07.03.2017

E-mail: [fao1976@ukr.net](mailto:fao1976@ukr.net), [immun56@gmail.com](mailto:immun56@gmail.com)  
 © А.А. Фефелов, В.И. Литвиненко, М.А. Таиф,  
 И.А. Лурье, 2017

UDC 574:004.9

А.А. Фефелов<sup>1</sup>, В.И. Литвиненко<sup>2</sup>, М.А. Таиф<sup>3</sup>, И.А. Лурье<sup>4</sup>

## Hybrid Approach for Gene Regulatory Networks Reconstruction by a System of Ordinary Differential Equations

<sup>1</sup> PhD in Techn. Sciences, Associate Professor, Department of Design of Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, [fao1976@ukr.net](mailto:fao1976@ukr.net)

<sup>2</sup> Doctor of Technical Sciences, Professor, Head of the Department of Informatics and Computer Science of Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, [immun56@gmail.com](mailto:immun56@gmail.com)

<sup>3</sup> Graduate student of the Department of Informatics and Computer Science, Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, [taifmohamedali@gmail.com](mailto:taifmohamedali@gmail.com)

<sup>4</sup> PhD in Techn. Sciences, Associate Professor, Department of Informatics and Computer Science of Kherson National Technical University, Bereslavskoe Shosse, 24, Kherson, 73008, Ukraine, [lurieira@gmail.com](mailto:lurieira@gmail.com)

**Keywords:** gene regulatory networks, reverse engineering, gene expression, ordinary differential equations, clonal selection algorithm, wavelet-neural network.

**Introduction.** Although there are a variety of models and methods for gene regulatory networks reconstruction, the problem of obtaining an adequate model based on experimental data is still urgent. In this regard, many studies use a fixed record of the differential equations based on the S-system. A significant disadvantage of such fixed record of the differential equations is the lack of flexibility of the model, what limits the scope of its application.

**Purpose.** The purpose of this work is development of the hybrid procedure of the solution of the gene networks reconstruction problem based on the ordinary differential equations.

**Method.** Models of the ordinary differential equations are used to model the gene regulatory networks. To solve the differential equations, wavelet-neural networks are used. The topology and tuning of the parameters is determined using the algorithm of the clonal selection. To find the concentration of gene expression products, which are represented by the method of solving the Cauchy problem, the Runge-Kutta method of the fourth order is applied.

**Results.** A hybrid method is developed that implemented the procedure for reconstructing gene regulatory networks based on the gene expression data. The effectiveness of the proposed method is proved by experimental studies that confirm the applicability of this approach to find the relationships between the components of the GRN.

**Conclusion.** The proposed work is a new Wavelet Neural Network and Clonal Algorithm approach for inferring Gene Regulatory Network which is expressed in terms of the ordinary differential model. The result of the proposed procedure is that further improvement of the technology, combined with preprocessing methods, will allow the effective reconstruction of real GRN. The main directions of further research we have chosen to create a meta-procedure for automatic configuration of parameters of a hybrid algorithm. This meta-procedure will reduce the search space by dynamically changing the intervals of representation of the elements values that make up the individuals of the AIS. Using this new method the bioinformatics and biologists can infer any Gene Regulatory Network of their interest. Also, they can understand the regulatory mechanism of the specific genes which causes the combat diseases.

