

УДК 574:004.2

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

Реконструкция S-системы гибридным алгоритмом клонального отбора и дифференциальной эволюции

Рассмотрено решение задачи реконструкции генной регуляторной сети, представленной в форме S-системы. Предложен гибридный алгоритм, основанный на методе клонального отбора. Проведены экспериментальные исследования влияния параметров гибридного алгоритма на качество решения задачи идентификации S-системы. На тестовых примерах проведен сравнительный анализ предложенного алгоритма с другими аналогичными вычислительными методами.

Ключевые слова: генные регуляторные сети, реверсная инженерия, экспрессия генов, обыкновенные дифференциальные уравнения, S-система, алгоритм клонального отбора, дифференциальная эволюция, структурно-параметрическая идентификация, сходимость алгоритма.

Розглянуто рішення задачі реконструкції генної регуляторної мережі, представлені у формі S-системи. Запропоновано гібридний алгоритм, заснований на методі клонального відбору. Проведено експериментальні дослідження впливу параметрів гібридного алгоритму на якість виконання задачі ідентифікації S-системи. На тестових прикладах проведено порівняльний аналіз запропонованого алгоритму з іншими аналогічними обчислювальними методами.

Ключові слова: генні регуляторні мережі, реверсна інженерія, реконструкція мереж, експресія генів, звичайні диференціальні рівняння, S-система, алгоритм клонального відбору, диференціальна еволюція, структурно-параметрична ідентифікація, збіжність алгоритму.

Введение. Реконструкция или реверсивная инженерия систем – это процесс вывода структурных и динамических характеристик изучаемой системы на основании наблюдений за ее поведением и определенных знаний в соответствующей предметной области. Сегодня реконструкция играет важнейшую роль в биологии, как один из основных инструментов моделирования биологических систем и их взаимодействий, имеющих ключевое значение для понимания механизмов их функционирования [1]. Одним из наиболее широко распространенных приложений методологии реконструкции в биологии есть идентификация генных регуляторных сетей (ГРС) [2]. Наличие и характер взаимодействий между генами ГРС принципиально для создания новых лекарственных препаратов. Реконструкция помогает исследователям найти ответы на ряд вопросов, среди которых можно выделить следующие:

- какие процессы в организме регулирует исследуемый ген;
- какие гены оказывают влияние на исследуемый ген;

- каким образом взаимодействуют гены;
- какие гены несут ответственность за тот или иной вид заболеваний;
- какие лекарственные препараты будут иметь эффект их применения в случае того или иного заболевания и т. д.

Технология ДНК-микрочипов [3] значительно облегчает изучение поведения ГРС, но задача реконструкции остается трудноразрешимой, так как содержит большое количество неизвестных.

Существует ряд моделей и методов реконструкции ГРС. Наиболее популярны из них: булевы модели [4], в которых взаимодействие между генами задано в форме булевых операций; линейные модели [5], где используется линейная взвешенная сумма уровней экспрессии генов; системы дифференциальных уравнений [6] – нелинейные и наиболее точные модели; ассоциативные и байесовские сети [7–9], с использованием статистического подхода при описании связей между генами; модели нейронных сетей [10–14], где нейроны ассоциируются с отдельными генами; гибридные

модели [11], комбинирующие несколько вычислительных парадигм.

В статье в качестве модели ГРС выбрана классическая система обыкновенных дифференциальных уравнений (ОДУ) в форме S -системы, для идентификации которой предложен гибридный алгоритм клонального отбора и дифференциальной эволюции.

Постановка проблемы

Реконструкция ГРС включает в себя выбор математической модели и разработку метода ее идентификации. Все существующие модели ГРС обладают своими достоинствами и недостатками. К основным характеристикам моделей можно отнести их способность обрабатывать зашумленные данные, учитывать нелинейности генных взаимодействий, точно описывать характер, взаимодействий. Сюда также относится вычислительная сложность процедур, необходимых для идентификации модели. S -система – одна из наиболее популярных моделей, используемых для описания ГРС [12]. Она нелинейна и поэтому способна очень точно воспроизвести поведение генной сети. С другой стороны, ее структура позволяет однозначно и достаточно просто преобразовать математическую форму записи в топологию, традиционно представленную в виде взвешенного ориентированного графа. Вершины графа соответствуют генам сети, а дуги определяют связи между этими генами. Таким образом S -система представляет компромисс между точностью и математической гибкостью.

S -система – разновидность системы ОДУ следующего вида:

$$\frac{dx_i}{dt} = \alpha_i \prod_{j=1}^N x_j^{g_{ij}} - \beta_i \prod_{j=1}^N x_j^{h_{ij}}, \quad (1)$$

где N – количество генов) в генной регуляторной сети; $x_i(t)$ – переменная, выражающая изменение уровня экспрессии i -го гена во времени; показатели степени g_{ij} и h_{ij} обуславливают направление воздействия и силу связи гена x_j с геном x_i ; α_i, β_i – коэффициенты, принимающие неотрицательные значения. С учетом биохимической инженерии параметры α_i, β_i носят на-

звание констант скорости, а g_{ij}, h_{ij} называются кинетическими порядками.

Данный формализм удобен своей структурой, которая выражает динамический баланс между воздействиями, приводящими к увеличению уровня экспрессии x_i (уменьшаемое), и воздействиями, приводящими к уменьшению уровня экспрессии x_i (вычитаемое). При этом положительные значения g_{ij} , к примеру, говорят об активирующем воздействии гена j на ген i . Графически это можно изобразить в виде дуги, направленной от j к i . Такая очевидная и простая графическая интерпретация обусловила высокую популярность S -системы.

Идентификация S -системы предполагает поиск оптимальных значений параметров из множества $v = \{\alpha, \beta, g, h\}$. Анализируя выражение (1) отметим, что данная задача не тривиальна ввиду высокой размерности. Для сети, состоящей из N генов, количество параметров, которые необходимо найти, определяется выражением $2N(N+1)$. Высокая вычислительная сложность задачи не позволяет решать ее аналитическими методами. В связи с этим разработка эффективного метода идентификации S -системы – актуальна.

Искусственные иммунные системы

Современные исследования в области иммунологии позволили сформировать особый взгляд на естественную иммунную систему животных и человека и рассматривать ее как мощную вычислительную структуру. Согласно этим исследованиям естественные иммунные механизмы высших существ обладают признаками, свойственными системам распознавания образов. Они способны обучаться распознавать чужеродные агенты, наносящие вред организму. Иммунные механизмы робастны к небольшим мутациям распознаваемых антигенов. Дальнейшее развитие иммунной вычислительной парадигмы привело к созданию целого ряда методов под общим названием – искусственные иммунные системы (ИИС) [13, 14]. В ИИС популяция индивидуумов (вариантов решений) из поколения в поколение подвергает-

ся воздействию иммунных операторов. В общем случае эти операторы – мутация и селекция. Популяция решений, первоначально генерируемая случайным образом, под воздействием мутации и селекции, в каждом новом поколении улучшает свои характеристики. Оценка популяции в количественном выражении проводится посредством присвоения индивидуумам значений аффинности. На основании оценок индивидуумы могут покидать популяцию или переходить в новое поколение. Способности к обучению и запоминанию сделали иммунную систему пригодной для решения задач оптимизации.

В работе [15] иммунная система рассматривается с учетом механизма клональной селекции. На основании принципа клональной селекции в [16] предложен алгоритм оптимизации *CLONALG*, широко используемый в настоящее время как одна из разновидностей ИИС. В клональном алгоритме значения аффинности выражают меру близости индивидуума к оптимальному решению и вычисляются на основании целевой функции задачи. Пошаговое описание алгоритма представлено ниже.

Шаг 1. Поколение = 0.

Шаг 2. Случайным образом создать начальную популяцию решений (Ab^0).

Шаг 3. Провести оценку популяции Ab^0 на основании целевой функции f .

Цикл пока условие останова $e = false$.

Шаг 4. Сохранить лучшее решение в текущем поколении.

Шаг 5. Выбрать антитела из Ab^0 с наибольшей аффинностью.

Шаг 6. Создать клоны Ab^c выбранных антител в количестве $n \sim f$.

Шаг 7. Провести мутацию клонов с интенсивностью $p_m! \sim f$.

Шаг 8. Провести оценку популяции Ab^c на основании целевой функции f .

Шаг 9. Выбрать клоны с наибольшей аффинностью из Ab^c и перенести их в Ab^0 .

Шаг 10. Заменить d худших антител в Ab^0 новыми случайно созданными.

Шаг 11. Провести оценку новых антител в Ab^0 .

Шаг 12. Поколение = Поколение + 1.

Конец цикла пока

Шаг 13. Вывод: лучшее решение в текущем поколении.

В *CLONALG*, в зависимости от типа задачи, можно использовать различные способы представления решений. Наиболее часто используются бинарное и вещественное представления. Также условия и цели задачи есть 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 данном этапе) входят во взаимодействие с антигеном, т.е. образуют минимумы целевой функции. В процессе активации отобранные антитела увеличивают свое представительство в пространстве решений путем клонирования. Клетки, чья аффинность выше, создают большее количество клонов, но меньше подвержены мутации. Мутация в *CLONALG* имеет высокую интенсивность, так как это главная движущая сила эволюции. В процессе замены, клетки с низкой аффинностью удаляются из основной популяции, на их место приходят новые случайно сгенерированные индивидуумы. Это позволяет избегать локальных экстремумов и исследовать всю целевую поверхность.

Сравнительные исследования производительности классического клонального алгоритма, проведенные авторами в [17], показали, что в контексте решения задачи реконструкции ГРС, *CLONALG* обладает недостатком, связанным с низкой эффективностью оператора мутации. В клональном алгоритме используется простая мутация, которая случайным образом (с заданной вероятностью) изменяет значение каждого элемента строки индивидуума:

$$Ab_{ij} = \begin{cases} \text{rand Init} (), & \text{если } \text{rand Event}(p_m) = 1 \\ Ab_{ij}, & \text{в противном случае} \end{cases}, j = 1, 2, \dots, L, \quad (2)$$

где Ab_{ij} – элемент строки i -го индивидуума популяции в бинарном или вещественном

представлении; L – длина строки индивидуума; p_m – интенсивность мутации; $\text{rand Init}()$ – функция случайной инициализации элемента строки индивидуума; $\text{rand Event}()$ – бинарная функция генерации случайного события с заданной вероятностью.

При увеличении количества параметров задачи растет L , что приводит к ухудшению работы мутации. В [17] эта проблема частично решается при помощи предложенной процедуры трансформации пространства решений. В статье предлагается способ повышения эффективности оператора мутации путем гибридизации с методом дифференциальной эволюции.

Дифференциальная эволюция

Алгоритм дифференциальной эволюции (ДЭ) – разновидность эволюционных алгоритмов [18, 19]. Обладая высокой эффективностью, ДЭ нашла применение во многих предметных областях, как метод глобальной оптимизации. Существует несколько вариантов ДЭ, отличающихся деталями реализации эволюционных операторов. В статье используется вариант, представленный в [18]. Краткое описание метода ДЭ приведено далее.

Рассматривается задача минимизации целевой функции:

$$f(x) \rightarrow \min, x = (x_1, \dots, x_n), \quad (3)$$

где x – вектор параметров задачи, на основании которого строятся индивидуумы популяции решений $x_i^G, i = 1, \dots, P$; P – размер популяции решений; G – текущее поколение.

Основное отличие алгоритма ДЭ от других эволюционных алгоритмов заключается в реализации оператора мутации. ДЭ-мутация выглядит следующим образом:

$$v_i^{G+1} = x_{r_3}^G + F(x_{r_1}^G - x_{r_2}^G), \quad (4)$$

где $v_i^{G+1}, i = 1, \dots, P$ – индивидуум, полученный в результате мутации; $r_1, r_2, r_3 \in [1, \dots, P]$ – индексы индивидуумов, которые выбираются случайно из популяции решений в текущем поколении такие, что $r_1 \neq r_2 \neq r_3 \neq i$; F – масштабный коэффициент $F \geq 0$.

Компоненты индивидуумов x_i^G частично заменяются соответствующими компонентами

векторов v_i^{G+1} с образованием популяции кандидатов $u_i^{G+1} = (u_{i1}^{G+1}, \dots, u_{in}^{G+1})$. Формирование вектора u_i^{G+1} происходит по следующему выражению:

$$u_{ij}^{G+1} = \begin{cases} v_{ij}^{G+1}, & \text{если } \text{rand Event}(p_{DE}) = 1 \vee j = k \\ x_{ij}^{G+1}, & \text{в противном случае} \end{cases}, j = 1, 2, \dots, n, \quad (5)$$

где $k \in \{1, \dots, n\}$ – случайный индекс параметра, выбираемый однажды для каждого индивидуума, смысл которого в том, чтобы гарантировать переход хотя бы одного компонента вектора v_i^{G+1} в вектор u_i^{G+1} ; p_{DE} – вероятность перехода j -го компонента вектора v_i^{G+1} в вектор u_i^{G+1} . Так как выражение (5) ассоциируется с оператором кроссинговера в эволюционных алгоритмах, p_{DE} называют вероятностью кроссинговера. Отметим, что выражение (5) очень похоже на выражение (2), поэтому p_{DE} названа интенсивностью ДЭ-мутации.

Популяция следующего поколения формируется из популяции текущего поколения и популяции кандидатов посредством ДЭ-селекции:

$$x_i^{G+1} = \begin{cases} u_i^{G+1}, & \text{если } f(u_i^{G+1}) \leq f(x_i^G) \\ x_i^G, & \text{в противном случае} \end{cases}, \quad (6)$$

т.е. каждый индивидуум из популяции кандидатов сравнивается с соответствующим индивидуумом из текущей популяции. Если кандидат обладает меньшим значением целевой функции, он переходит в новое поколение. Иначе в новое поколение переходит текущий индивидуум.

Предлагаемый гибридный алгоритм

Идея объединения различных вычислительных методов и образования гибридов основывается на предположении о том, что полученный в результате объединения новый вычислительный метод должен обладать более высокой производительностью, чем входящие в него 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, участвующих 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 статье предложен гибридный клональный алгоритм, в котором фаза мутации расширена операторами, взятыми из алгоритма дифференциальной эволюции. Пошаговое описание алгоритма представлено ниже.

Шаг 1. Поколение = 0.

Шаг 2. Случайным образом создать начальную популяцию решений (Ab^0).

Шаг 3. Провести оценку популяции Ab^0 на основании целевой функции f .

Цикл пока условие останова $e = false$.

Шаг 4. Сохранить лучшее решение в текущем поколении.

Шаг 5. Выбрать антитела из Ab^0 с наибольшей аффинностью.

Шаг 6. Создать клоны Ab^c выбранных антител в количестве $n \sim f$.

Шаг 7. Провести мутацию клонов с вероятностью $p_c! \sim f$ и интенсивностью $p_m! \sim f$ по формуле (2).

Шаг 8. Провести ДЭ-мутацию клонов с интенсивностью $p_{DE}! \sim f$ по формулам (4) и (5) и записать результаты в популяцию кандидатов Ab^t .

Шаг 9. Провести оценку популяции Ab^t на основании целевой функции f .

Шаг 10. Провести ДЭ-селекцию кандидатов по формуле (6).

Шаг 11. Выбрать кандидаты с наибольшей аффинностью из Ab^t и перенести их в Ab^0 .

Шаг 12. Заменить d худших антител в Ab^0 новыми случайно созданными.

Шаг 13. Провести оценку новых антител в Ab^0 .

Шаг 14. Поколение = Поколение + 1.

Конец цикла пока

Шаг 15. Вывод: лучшее решение в текущем поколении.

Шаг 7 оставлен для того, чтобы предотвратить преждевременную сходимость алгоритма. В отличие от классического *CLONALG* в данном случае мутации подвергаются не все клоны, а лишь часть из них, что управляется параметром p_c . При этом значения вероятности p_c и интенсивности p_m – предмет исследования. Кроме этого, авторами проведены исследования влияния интенсивности ДЭ-мутации p_{DE} , масштабного коэффициента F и ДЭ-селекции на качество работы алгоритма, результаты которых представлены далее.

В предложенном алгоритме реализовано вещественное кодирование антител, проводимое по схеме, изображенной на рис. 1.

Для расчета аффинности индивидуумов выбрана относительная квадратическая ошибка аппроксимации данных временных рядов экспрессии генов, представленная следующим выражением [20]:

$$f = \text{MIN} \sum_{i=1}^N \sum_{j=1}^T \left(\frac{x_i^M(t_0 + j\Delta t) - x_i(t_0 + j\Delta t)}{x_i(t_0 + j\Delta t)} \right)^2, \quad (7)$$

где t_0 – момент времени начала измерения концентрации продуктов экспрессии; Δt – интервал времени между последовательными измерениями концентрации продуктов экспрессии; T – количество измерений; $x_i^M(t_0 + j\Delta t)$ – значения концентрации, полученные моделированием, т.е. решением системы (1); $x_i(t_0 + j\Delta t)$ – наблюдаемые значения концентрации, полученные при помощи измерений.

Значения концентрации x_i^M получены решением системы ОДУ (1) одним из методов численного интегрирования (в данном случае методом Рунге–Кутты) с начальными условиями $x_i(t_0)$.

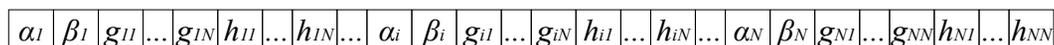


Рис. 1. Структура антитела, кодирующего S-систему

Исследования

Эксперименты включают исследования, состоящие из двух этапов. На первом этапе рассматривается задача реконструкции искусственной ГРС, состоящей из двух генов [20]. Реконструкция проводится для сети с фиксированной и с нефиксированной структурой. Фиксированная структура сети в отличие от нефиксированной предполагает наличие информации о связях между генами (т.е. матрица связей считается константой), что упрощает решение, так как алгоритм занимается поиском только весов существующих связей. На втором этапе рассматривается ГРС, состоящая из четырех генов [21]. В данном случае исследуется сеть с фиксированной структурой. Для сравнения результатов исследований выбран классический [16] и модифицированный [22] алгоритм клонального отбора.

Общие параметры алгоритма реконструкции, которые не меняются на протяжении всех экспериментов, приведены в табл. 1.

Таблица 1. Параметры алгоритма реконструкции для всех экспериментов

Название параметра	Значение параметра
Способ кодирования индивидуумов	Вещественный
Размер основной популяции антител	300
Размер популяции клонов	900
Коэффициент отбора	1,0 ($n = 300$)
Тип отбора	Турнирный
Размер турнира	4

Значение коэффициента отбора, равное 1,0, автоматически отключает шаги 12 и 13 гибридного алгоритма, т.е. *подпитка* новыми случайными индивидуумами не проводится. Такая установка позволяет получить более точные результаты исследования влияния параметров p_c и p_m , так как функция шага 7 в разработанном методе очень похожа на функцию шагов 12 и 13.

Для проведения экспериментов использовалась информационная система реконструкции ГРС, объектно-ориентированная архитектура которой представлена в [22].

Реконструкция ГРС, состоящей из двух генов

В данном эксперименте исследовано влияние параметров p_{DE} , p_c , p_m и оператора ДЭ-се-

лекции на качество работы гибридного алгоритма, где используется фиксированная структура ГРС.

Параметры S -системы для двухгенной ГРС ($N = 2$) представлены в табл. 2. При построении временного ряда наблюдаемых данных экспрессии, необходимых для проведения экспериментов, система (1) решена с настройками, взятыми из табл. 2. Решение S -системы осуществлялось методом Рунге–Кутты четвертого порядка с количеством временных отсчетов $T = 50$ и начальными условиями $(x_1(t_0), x_2(t_0)) = (1,0; 1,5)$.

Таблица 2. Параметры S -системы ГРС, состоящей из двух генов

i	α_i	β_i	g_{i1}	g_{i2}	h_{i1}	h_{i2}
1	3,0	3,0	0,0	-2,5	-1,0	0,0
2	3,0	3,0	2,5	0,0	0,0	2,0

Параметры алгоритма реконструкции, используемые в данном эксперименте, представлены в табл. 3. В первом случае авторы исследовали влияние оператора ДЭ-селекции, поэтому значения остальных исследуемых параметров зафиксированы.

Таблица 3. Параметры алгоритма реконструкции в эксперименте с двухгенной ГРС

Название параметра	Значение параметра
Начальные промежутки значений для параметров α и β	[0,0; 5,0]
Начальные промежутки значений для параметров g и h	[-3,0; 3,0]
Количество поколений (условие останова e)	1500
Вероятность простой мутации p_c	0,1
Интенсивность простой мутации p_m	0,05
Интенсивность ДЭ-мутации p_{DE}	0,5
Масштабный коэффициент F	0,8

Для проверки влияния ДЭ-селекции авторы провели по 10 пусков алгоритма с участием и без участия ДЭ-селекции, а результаты усреднили и свели в табл. 4.

Результат позволяет сделать однозначный вывод и отказаться от ДЭ-селекции во всех последующих экспериментах.

Таблица 4. Сравнительная таблица ошибок модели при исследовании влияния оператора ДЭ-селекции

Ошибка модели (f)	ДЭ-селекция включена	ДЭ-селекция выключена
Минимальная	0,00024	$6,87 \cdot 10^{-10}$
Максимальная	0,0572	$1,31 \cdot 10^{-6}$
Средняя	0,0149	$3,2 \cdot 10^{-7}$

Влияние масштабного коэффициента исследовалось на интервале значений параметра $[0,2; 1,4]$, разделенного на 13 отсчетов. Для каждого отсчета проведено по 10 пусков алгоритма с последующим усреднением результатов (рис. 2).

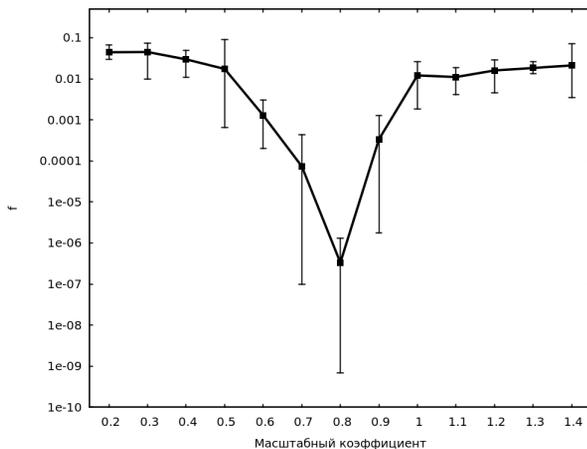


Рис. 2. Диаграмма ошибок модели при исследовании влияния множителя F

Диаграмма, представленная на рис. 2, состоит из средних ошибок модели (жирная линия) и двух полочек в каждом отсчете. Полочки соответствуют максимальному (верхняя полочка) и минимальному (нижняя полочка) значению ошибки модели. Как видно из рисунка, результат $F = 0,8$ оказался оптимальным, так как минимизирует значение ошибки.

Для проверки влияния интенсивности ДЭ-мутации интервал значений данного параметра $[0,01; 1,0]$ разделили на 11 отсчетов с 10 пусками алгоритма на каждый отсчет и последующим усреднением результатов (рис. 3).

Результат, свидетельствующий о том, что операция кроссинговера, которая проводится по выражению (5), в данном случае несет скорее негативный, чем позитивный эффект. Тем не менее авторы продолжили эксперименты с фиксированным значением $p_{DE} = 0,5$.

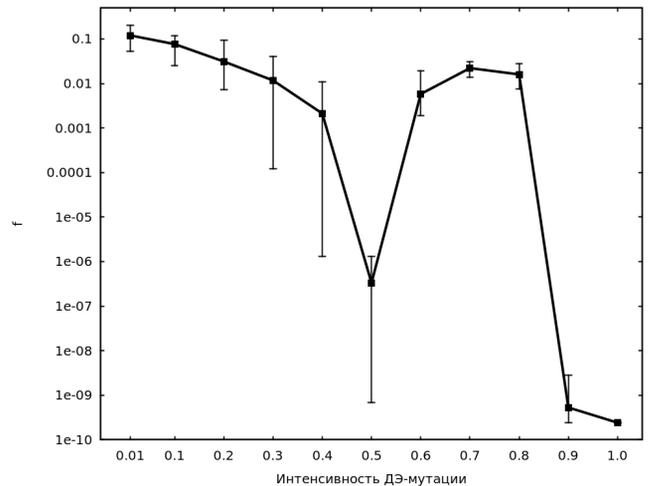


Рис. 3. Диаграмма ошибок модели при исследовании влияния параметра p_{DE}

На рис. 4 и 5 показаны результаты исследования влияния вероятности и интенсивности простой мутации. Для каждого из параметров выбрано по 12 отсчетов в интервалах, соответственно $[0,0; 1,0]$ и $[0,01; 1,0]$. Как и в предыдущих исследованиях, здесь для каждого отсчета проведены по 10 пусков алгоритма реконструкции.

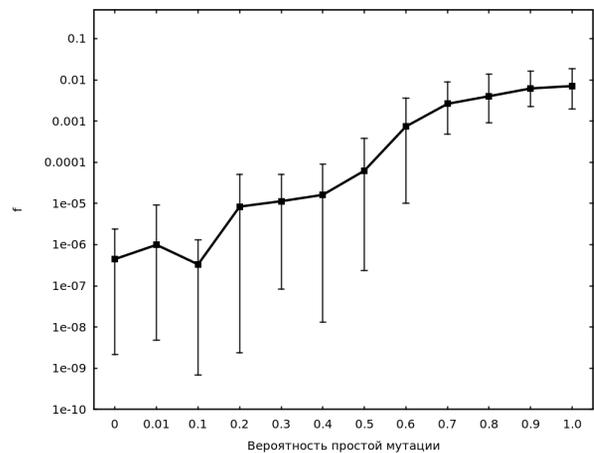


Рис. 4. Диаграмма ошибок модели при исследовании влияния параметра p_c

Результаты говорят о том, что простая мутация в исследуемом алгоритме должна быть достаточно низкой как по вероятности, так и по интенсивности, но не нулевой. В более сложных задачах отсутствие простой мутации будет приводить к преждевременной сходимости алгоритма и невозможности отыскания глобального оптимума.

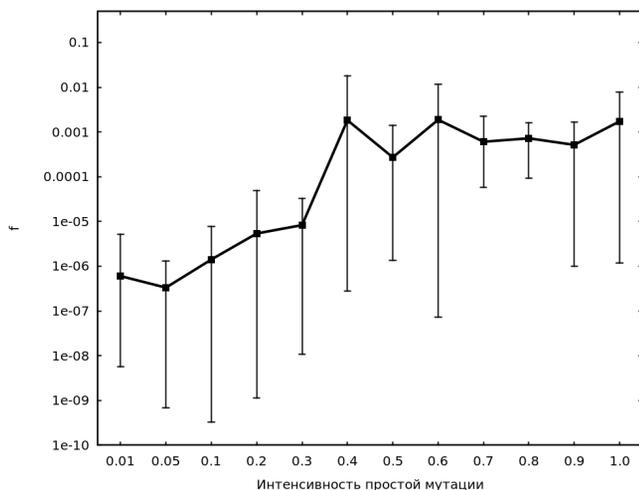


Рис. 5. Диаграмма ошибок модели при исследовании влияния параметра p_m

Сравнение результатов работы трех ИИС, основанных на клональной селекции, применительно к задаче реконструкции двухгенной ГРС, представлено в табл. 5.

Таблица 5. Сравнительная таблица ошибок модели при исследовании трех методов, основанных на клональной селекции

Ошибка модели (f)	Исследуемый гибридный алгоритм	Модифицированный клональный алгоритм [17]	Классический клональный алгоритм [16]
Минимальная	$6,87 \cdot 10^{-10}$	$4,275 \cdot 10^{-7}$	0,0027
Максимальная	$1,31 \cdot 10^{-6}$	0,0035	0,541
Средняя	$3,32 \cdot 10^{-7}$	$6,0 \cdot 10^{-4}$	0,154

Авторы также провели исследование качества реконструкции двухгенной ГРС при нефиксированной структуре. При этом операция кроссинговера исключалась из алгоритма, т.е. $p_{DE} = 1,0$. В результате всех 10 пусков алгоритм выдавал минимально возможное (ограничения связаны с программной реализацией) значение ошибки модели, равное $2,36502 \cdot 10^{-10}$. Сравнение вычисленных параметров S -системы с аналогичными результатами, представленными в работе [20], показано в табл. 6.

Испытания проводились с использованием аппаратного обеспечения *Intel Core i3-3110M* (2.4 ГГц). При этом было задействовано только одно ядро процессора. В последнем испытании среднее время работы гибридного алгоритма до получения минимальной ошибки составило 53 секунды.

Таблица 6. Сравнительная таблица значений параметров S -системы в эксперименте с нефиксированной структурой ГРС

i	α_i	β_i	g_{i1}	g_{i2}	h_{i1}	h_{i2}
Исследуемый гибридный алгоритм						
1	3,0003	3,0003	0,00	-2,4998	-0,9999	-0,0001
2	3,0001	3,0001	2,4999	0,00	0,00	1,9999
Алгоритм реконструкции, описанный в [20]						
1	3,71	3,74	0,00	-1,91	-0,77	0,00
2	2,93	3,00	2,58	0,00	0,00	1,88

График сходимости алгоритма в последнем испытании показан на рис. 6.

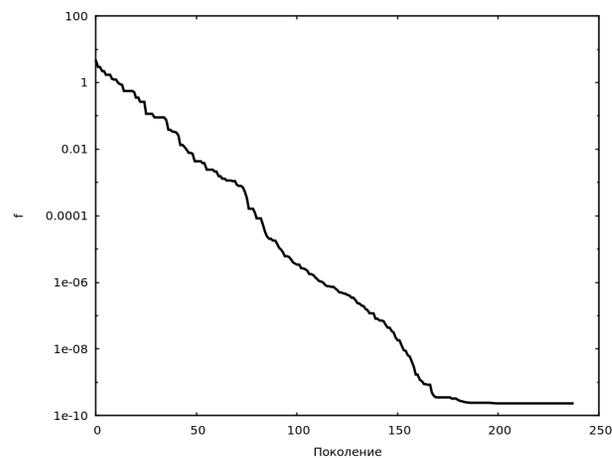


Рис. 6. График сходимости предложенного алгоритма при реконструкции двухгенной ГРС с нефиксированной структурой (в логарифмическом масштабе)

Реконструкция ГРС, состоящей из четырех генов

На данном этапе авторы увеличили сложность задачи, выбрав регуляторную сеть, состоящую из четырех генов ($N = 4$). В табл. 7 приведены параметры ГРС, используемые для генерирования временных рядов данных экспрессии генов. Временные ряды генерируются решением S -системы (1) с начальными условиями $(x_1(t_0), x_2(t_0), x_3(t_0), x_4(t_0)) = (1,4; 2,7; 1,2; 0,4)$ и количеством измерений $T = 26$.

Таблица 7. Параметры S -системы для ГРС, состоящей из четырех генов

i	α_i	β_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	h_{i1}	h_{i2}	h_{i3}	h_{i4}
1	12,0	10,0	0,0	0,0	-0,8	0,0	0,5	0,0	0,0	0,0
2	8,0	3,0	0,5	0,0	0,0	0,0	0,0	0,75	0,0	0,0
3	3,0	5,0	0,0	0,75	0,0	0,0	0,0	0,0	0,5	0,2
4	2,0	6,0	0,5	0,0	0,0	0,0	0,0	0,0	0,0	0,8

Параметры алгоритма реконструкции, используемые в данном эксперименте, представлены в табл. 8. Исследование проводилось при фиксированной структуре ГРС. Результаты усреднялись на основании 10 пусков алгоритма.

Таблица 8. Параметры алгоритма реконструкции в эксперименте с четырехгенной ГРС

Название параметра	Значение параметра
Начальные промежутки значений для параметров α и β	[0,0, 15,0]
Начальные промежутки значений для параметров g и h	[-1,0, 1,0]
Количество поколений (условие останова e)	3000
Вероятность простой мутации p_c	0,1
Интенсивность простой мутации p_m	0,05
Интенсивность ДЭ-мутации p_{DE}	1,0
Масштабный коэффициент F	0,8

Сравнение результатов работы трех ИИС, основанных на клональной селекции, применительно к задаче реконструкции ГРС, состоящей из четырех генов, представлено в табл. 9.

Следует отметить, что в данном испытании из 10 пусков в 60 процентах случаев алгоритм сходился к минимально возможной ошибке аппроксимации $1,634 \cdot 10^{-10}$, а в 40 процентах случаев – к ошибке 0,00068, после чего начиналась стагнация. Стагнация – явление, свойственное дифференциальной эволюции [23]. Оно требует проведения дополнительных исследований, которые выходят за рамки статьи.

Вычисленные параметры S -системы в испытании с минимальной ошибкой модели, показаны в табл. 10.

Как видно из таблицы, рассчитанные значения практически совпадают с оригинальными значениями в табл. 7. График сходимости алгоритма в испытании с лучшим результатом показан на рис. 7.

Таблица 10. Рассчитанные параметры S -системы для ГРС, состоящей из четырех генов

i	α_i	β_i	g_{i1}	g_{i2}	g_{i3}	g_{i4}	h_{i1}	h_{i2}	h_{i3}	h_{i4}
1	11,9999	9,9999	0,0	0,0	-0,80002	0,0	0,50001	0,0	0,0	0,0
2	7,9997	2,9997	0,50003	0,0	0,0	0,0	0,0	0,75005	0,0	0,0
3	2,9987	4,9996	0,0	0,75027	0,0	0,0	0,0	0,0	0,50028	0,20021
4	1,9999	6,0001	0,50002	0,0	0,0	0,0	0,0	0,0	0,0	0,80003

Таблица 9. Сравнительная таблица ошибок модели при исследовании трех методов, основанных на клональной селекции

Ошибка модели (f)	Исследуемый гибридный алгоритм	Модифицированный клональный алгоритм [17]	Классический клональный алгоритм [16]
Минимальная	$1,634 \cdot 10^{-10}$	0,00112	0,03
Максимальная	0,00068	0,0089	1,021
Средняя	0,000272	0,00466	0,6

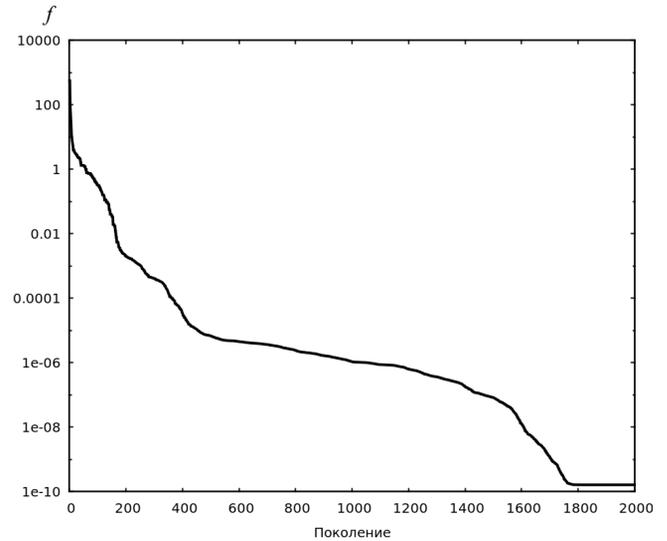


Рис. 7. График сходимости предложенного алгоритма при реконструкции ГРС, состоящей из четырех генов (в логарифмическом масштабе)

Заключение. Предложен гибридный метод реконструкции ГРС, позволяющий увеличить скорость сходимости и точность алгоритма оптимизации при решении задачи идентификации S -системы. В основе метода лежит технология гибридизации, объединяющая искусственную иммунную систему в форме алгоритма клонального отбора, и алгоритм дифференциальной эволюции. Для проверки эффективности предложенного подхода проведены эксперименты по реконструкции искусственных ГРС. Во время экспериментов исследовалось влияние различных параметров гибридного алгоритма на качество решения задачи рекон-

струкции S-системы, а также проводилось сравнение с другими вычислительными методами. Эксперименты показали негативный эффект от применения операторов дифференциальной эволюции, таких как селекция и кроссинговер. С другой стороны, существенный положительный эффект показал оператор ДЭ-мутации. Результаты сравнительных экспериментов подтвердили преимущества разработанного гибридного метода и алгоритма перед аналогичными вычислительными методами.

1. *Csete M., Doyle J.* Reverse Engineering of Biological Complexity // *Science*. – 2002. – **295**. – P. 1664–1669.
2. *Reverse engineering of gene regulatory networks / K.H. Cho, S.M. Choo, S.H. Jung et al.* // *IET System Biology*. – 2007. – **1**. – P. 149–163.
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. *Liang S., Fuhrman S., Somogyi R.* Reveal, a general reverse engineering algorithm for inference of genetic network architectures // *Pacific Symposium on Biocomputing (PSB '98)*, Jan. 1998: proceedings. – Maui, Hawaii, USA, 1998. – P. 18–29.
5. *Linear modeling of mRNA expression levels during CNS development and injury / P. D'haeseleer, X. Wen, S. Fuhrman et al.* // *4th Pacific Symposium on Biocomputing (PSB '99)*, Jan. 1999: *Ibid.*, 1999. – P. 41–52.
6. *Chen T., He H.L., Church G.M.* Modeling gene expression with differential equations // *Ibid.* – P. 29–40.
7. *Schafer J., Strimmer K.* An empirical Bayes approach to inferring large-scale gene association networks // *Bioinformatics*. – 2005. – **21**(6). – P. 754–764.
8. *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.
9. *Reconstruction and validation of gene regulatory networks with neural networks / H. Hache, C. Wierling, H. Lehrach et al.* // *2nd Foundations of Systems Biology in Engineering Conference (FOSBE '07)*, Sept. 2007: proceedings. – Stuttgart, Germany, 2007. – P. 319–324.
10. *Hybrid Approach for Gene Regulatory Networks Reconstruction by a System of Ordinary Differential Equations / A.A. Fefelov V.I. Lytvynenko M.A. Taif et al.* // *Upr. sist. maš.* – 2017. – № 3 – P. 63–72. (In Russian).
11. *Modeling T-cell activation using gene expression profiling and state-space models / C. Rangel, J. Angus, Z. Ghahramani et al.* // *Bioinformatics*. – 2004. – **20**(9). – P. 1361–1372.
12. *Savageau M.A.* Introduction to S-systems and the underlying power-law formalism // *Mathematical and Computer Modelling*. – 1988. – **11**. – P. 546–551.
13. *Dasgupta D.* Artificial Immune Systems and Their Applications. – Springer-Verlag, 1999. – 306 p.
14. *De Castro L.N., Timmis J.* Artificial Immune Systems: A New Computational Intelligence Approach. – Heidelberg, Springer, 2002. – 357 p.
15. *Burnet F.M.* A modification of jerne's theory of antibody production using the concept of clonal selection // *CA: a cancer journal for clinicians*. – 1976. – **26**(2). – P. 119–121.
16. *De Castro L.N., Zuben Von. F.J.* Learning and optimization using the clonal selection principle // *IEEE Transactions on Evolutionary Computation*. – 2002. – **6**(3). – P. 239–251.
17. *Parametric Identification of the S-System by the Modified Clonal Selection Algorithm / A.A. Fefelov V.I. Lytvynenko M.A. Taif et al.* // *Upr. sist. maš.* – 2017. – № 5 – P. 43–53. (In Russian).
18. *Storn R., Price K.V.* Differential evolution — a simple and efficient heuristic for global optimization over continuous spaces // *J. of Global Optimization*. – 1997. – **11**(4). – P. 341–359.
19. *Storn R., Price K.V.* Minimizing the real function of the ICEC'96 contest by differential evolution // *IEEE Int. Conf. on Evolutionary Computation*, May 1996: proceedings. – Nagoya, Japan, 1996. – P. 842–844.
20. *Tominaga D., Koga N., Okamoto M.* Efficient numerical optimization algorithm based on genetic algorithm for inverse problem // *Genetic and Evolutionary Computation Conference (GECCO '00)*, July 2000: proceedings. – Las Vegas, Nevada, USA, 2000. – **251**. – P. 251–258.
21. *Kim K.-Y., Cho D.-Y., Zhang B.-T.* Multi-stage evolutionary algorithms for efficient identification of gene regulatory networks // *Applications of evolutionary computing: EvoWorkshops 2006*, Apr. 2006: proceedings. – Budapest, Hungary, 2006. – P. 45–56.
22. *Object-Oriented Architecture of the Information System for the Reconstruction of the Gene Regulatory Networks / A.A. Fefelov V.I. Lytvynenko M.A. Taif et al.* // *Upr. sist. maš.* – 2017. – № 4 – P. 67–75, 82. (In Russian).
23. *Lampinen J., Zelinka J.* On stagnation of the differential evolution algorithm // *6th Int. Conf. on Soft Computing MENDEL 2000*, June 2000: proceedings. – Brno, Czech Republic, 2000. – P. 76–83.

Поступила 04.10.2017

E-mail: fao1976@ukr.net, immun56@gmail.com,
taifmohamedali@gmail.com, mary_voronenko@i.ua

© А.А. Фефелов, В.И. Литвиненко, М.А. Таиф,
М.А. Вороненко, 2017

Andrej Fefelov¹, Volodymyr Lytvynenko², Muchamed Ali Taif³, Mariia Voronenko⁴

¹ PhD in Techn.Sciences, Associated Professor, the Design Department of Kherson National Technical University, Beryslavske highway, 24, Kherson, 73008, Ukraine, fao1976@ukr.net

² Doctor of Technical Sciences, Professor, Head of the Department of Informatics and Computer Science of Kherson National Technical University, Beryslavske highway, Kherson, 73008, Ukraine, immun56@gmail.com

³ Graduate student of the Department of Informatics and Computer Science, Kherson National Technical University, Beryslavske highway 24, Kherson, 73008, Ukraine, taifmohamedali@gmail.com

⁴ PhD in Techn. Sciences, Associate Professor, the Department of Informatics and Computer Science of Kherson National Technical University, Beryslavske highway, 24, Kherson, 73008, Ukraine, mary_voronenko@i.ua

Reconstruction of the S-System by the Hybrid Algorithm of the Clonal Selection and Differential Evolution

Keywords: gene regulatory networks, reverse engineering, gene expression, ordinary differential equations, S-system, clonal selection algorithm, differential evolution, structural-parametric identification, convergence of the algorithm.

Introduction. Reconstruction or reverse engineering is a process of a conclusion of the structural and dynamic characteristics of the system under study on the basis of the observations of its behavior and its specific knowledge in the subject domain. Now days, many different models and methods of gene regulatory reconstruction have been developed, which have both advantages and disadvantages. At the choice of descriptive model it is necessary to consider the fact that the mathematical models, as a rule, have their own structure and a number of parameters which need to be identified. Different algorithms have been used so far to address the problem of finding the gene regulatory network. Most of them are evolutionary approaches or swarm based approaches. Genetic algorithms, genetic programming, memetic algorithm, clonal algorithm, membrane algorithms, evolution strategy, differential evolution, self adaptive differential evolution, particle swarm optimization are some of the widely used techniques. These algorithms are population-based search algorithms. Most of them starts with the random solutions and try to improve it using the multiple iterations.

Purpose. The aim of this work is to develop an effective hybrid method for the reconstruction of the gene regulatory networks, which will increase the rate of convergence in solving the problem of the S-system optimizing.

Method. We propose a hybrid method for reconstructing the GRN. This method is based on the hybridization technology, which allows combining the best qualities of the algorithm of clonal selection and the algorithm of differential evolution.

Results. We propose a hybrid method of reconstruction GRN, allowing to increase the convergence rate and accuracy of the optimization algorithm to solve the problem of identification S-system. The S-system was applied, as a computational model. Parameters and structure were calculated by using the clonal selection algorithm and algorithm differential evolution. The gene expression profiles are used as input data. They were represented by time series of changes in the expression products concentration. The experiments have shown the negative effect of the differential evolution operators application such as selection and crossover. On the other hand, a significant positive effect of mutation operator is shown, which is used in the algorithm of the differential evolution.

Conclusion. The comparative experiments results confirmed the advantages developed a hybrid method and an algorithm similar to the computational methods. The developed method and algorithm allowed to increase the speed of algorithms optimization convergence and simultaneously to increase their accuracy in solving the problem of the S-system reconstruction. It can be used to modify evolutionary algorithms or artificial immune systems.



Внимание !

**Оформление подписки для желающих
опубликовать статьи в нашем журнале обязательно.**

В розничную продажу журнал не поступает.

Подписной индекс 71008