



## ПРОБЛЕМНЫЕ СТАТЬИ

---

---

БИТ

*М. Л. Гаранина, К. Я. Кудрявцев*

### ПРИМЕНЕНИЕ МЕХАНИЗМА ГЕНОВ-ПЕРЕКЛЮЧАТЕЛЕЙ ДЛЯ СОЗДАНИЯ ЭФФЕКТИВНЫХ СИСТЕМ ОБЕСПЕЧЕНИЯ БЕЗОПАСНОСТИ ДАННЫХ НА ОСНОВЕ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Сегодня никто не будет спорить с тем, что информационные технологии заняли в нашей жизни существенное место, изменили ее и продолжают свое проникновение в нее на всех уровнях. К основным изменениям можно отнести повсеместный переход к хранению информации в электронном виде, широкую доступность средств фиксации событий (разнообразные фото- и видеокамеры, диктофоны) и развитие Интернета. Такое активное и, что более важно, качественное изменение информационных технологий заставляет по-новому взглянуть на проблему безопасности информации, причем на всех уровнях: личном, корпоративном, государственном [1].

В современных условиях меняться должны и технологии, касающиеся безопасности информации, и изменения эти должны быть быстрыми и эффективными. В статье будет рассмотрено обеспечение конфиденциальности частной, приватной информации обычных людей.

Легкое получение информации, ее мгновенное и фактически бесконтрольное распространение вкупе с практически вечной памятью Всемирной паутины приводит к тому, что современным пользователям приходится сталкиваться с проблемой размытия границ своей частной жизни и необходимостью определения уровней доступа к информации о себе для других [2]. Совершенно очевидно, что глупо отказываться от тех преимуществ, которые дают нам современные информационные технологии, и единственным решением оказывается определение того, кому и что можно знать о нас, а также *контроль над доступом* к этой информации.

Разумеется, в данном случае, как и в большинстве остальных, решение должно быть комплексным, а именно сочетать в себе как современные технологии защиты информации, так и законодательную базу, а также формирование соответствующего уровня грамотности пользователей (кто из нас не пересылал на домашнюю почту конфиденциальные документы, чтобы с ними спокойно поработать?).

В данной статье предлагается более подробно остановиться на требованиях к современным технологиям защиты информации, объективным и субъективным. Начнем с тех, которые порождаются объективными причинами, и прежде всего высокой скоростью изменения угроз. На первый план здесь выходят гибкость и эффективность реакции на новые типы угроз, а также доступная цена программного средства для большого количества пользователей. Что касается субъективных требований, то здесь важное значение имеет возможность контроля уровня доступа к той или иной личной информации, что позволит каждому самостоятельно устанавливать границы

своей частной жизни. Таким образом, возможность самостоятельной работы (без дополнительного участия пользователя), адаптация к новым требованиям потребуются и здесь.

Анализируя сформулированные требования (адаптивность, самостоятельность, быстрота реакции на новые виды угроз, удобство для пользователя в плане минимальных временных затрат на работу с ПО), можно предположить, что больше остальных для решения данной задачи будут подходить системы, построенные на основе технологий искусственного интеллекта. Именно они обладают способностью в меняющихся условиях непрерывно адаптироваться и обучаться и, что еще более важно, могут делать это самостоятельно, без постоянного участия разработчика [3].

Однако здесь возникает препятствие в виде экономической стороны вопроса. Следовательно, появляется задача удешевления технологий по созданию систем искусственного интеллекта. Решением в данном случае мог бы стать так называемый коробочный вариант системы искусственного интеллекта, предназначенной для защиты частной информации и контроля над доступом к ней, в таком случае один из дорогостоящих этапов — настройку и обучение — можно было бы перенести на период начала эксплуатации ПО пользователем. Система может непрерывно обучаться, работая с данными пользователя в реальных условиях, при необходимости дополнительные обучающие и контрольные данные могут быть включены в комплект с программным обеспечением. Таким образом, в течение всего времени работы система, непрерывно эволюционируя, будет адаптироваться к изменениям внешней среды.

Но для создания такой «коробки» потребуются разработка новой технологии обучения и постоянной адаптации систем искусственного интеллекта.

В данной статье предлагается рассмотреть в качестве подхода к решению поставленной задачи эволюционные вычисления. Для использования их в разработке массового решения предлагается методика по применению усовершенствованных генетических алгоритмов для повышения эффективности процесса адаптации систем искусственного интеллекта. Предлагаемая методика подразумевает внесение ряда изменений в классические этапы метода генетических алгоритмов, которые позволят применить биологический механизм генов-переключателей для адаптации к новым условиям.

Прежде чем переходить к описанию предлагаемой методики по применению генов-переключателей, рассмотрим, как предлагаемый механизм работает в природе.

После того как генетика достигла значимых результатов в области расшифровки ДНК живых существ нашей планеты, возник интересный вопрос: если ДНК всех живых существ так схожи, то откуда же такое немыслимое разнообразие видов? Как объяснить громадную разницу между мышью и человеком, если их геном совпадает на 99 %, а с акулой у человека ДНК совпадает на 30 %? Каким образом природа из одного и того же материала получает столь разный результат? В течение 40 лет ученые изучали различные участки ДНК, кодирующие белки, которые, отвечают за те или иные внешние признаки у животных. Логично предположить, что если определенный ген есть, то есть и анатомический признак, который он кодирует, и наоборот.

Однако выяснилось, что исходный набор строительных материалов во многом одинаков. В ходе этих исследований также был найден особый механизм — генетические переключатели, помощью которого природа и эволюция управляют работой генов-свойств (гены, кодирующие свойства), т. е. определяют, где и когда они будут активированы. Первые такие переключатели были обнаружены в 2002 г. Рональдом Брейкером и его коллегами из Йельского университета. С тех пор число публикаций, посвященных этому странному и очень древнему механизму генной регуляции, стремительно растет.

В эволюции анатомических особенностей живых существ ключевую роль играет именно эволюция генов-переключателей, а не генов-свойств. Геном состоит из кодирующей и некодирующей

частей, у человека, например, на долю кодирующей части приходится всего 1,5 % всего генома. Функции большинства некодирующих областей в геноме на текущий момент мало изучены, однако известно, что одной из этих функций является управление экспрессией генов. Экспрессия гена — это транскрипция гена в матричную РНК (мРНК) и последующая трансляция мРНК в белок [4]. Так как клетке невыгодно тратить энергию на синтез ненужных ей мРНК и белков, то многие гены синтезируются только в определенных органах, тканях и клетках. И именно некодирующие участки определяют, где и когда должна произойти транскрипция каждого определенного гена. В данном случае используется принцип переключателя. В ДНК существует некодирующий участок, исполняющий роль генетического переключателя с двумя положениями: «включено» и «выключено».

Изменение состояния такого переключателя (энхансера) вызывается действием специальных белков — факторов транскрипции. При их присоединении к энхансерам положение переключателя переводится в состояние «включено», запускается транскрипция гена, связанного с данным энхансером. У каждого гена есть один или несколько независимых энхансеров, каждый из которых независимо от других отвечает за экспрессию данного гена в разных частях тела в разное время жизненного цикла организма. Перед нами потрясающий метод оптимизации, придуманный самой природой. Один и тот же ген может использоваться в разных местах, что существенно расширяет круг его функций.

В применении к процессу создания систем искусственного интеллекта (ИИ) основные положения методики генов-переключателей в целом заключаются во внесении ряда изменений в основные этапы применения генетических алгоритмов (генерация начальной популяции, кроссинговер, селекция, мутация), которые позволят реализовать возможности генов-энхансеров.

Методика предполагает, во-первых, внесение изменений в принцип формирования самого генома, затрагивая, таким образом, первый из указанных выше этапов, а во-вторых, включение в задачу оптимизации системы ИИ влияния предметной области, в которой она будет работать, и наших знаний о ней [5]. Рассмотрим суть методики подробнее.

### **1. Изучение и параметризация предметной области. Определение факторов транскрипции**

При разработке систем, использующих в своей работе технологии искусственного интеллекта, часто требуется учитывать влияние динамически изменяющихся факторов внешней среды. Для упрощения работ по параметризации предметной области предлагается разбить параметры внешней среды на определенные классы в зависимости от их свойств или особенностей [6].

Таким образом, задача сводится к нахождению набора факторов транскрипции, отражающих параметры предметной области, определению их ранга и области допустимых значений.

Для выявления всех параметров предметной области для определения набора факторов транскрипции, достаточных для эффективного применения механизма генов-переключателей при разработке систем искусственного интеллекта, наиболее оправданным является применение методов инженерии знаний. Для выявления факторов транскрипции наиболее подходящими среди основных стратегий получения знаний являются формирование и извлечение знаний [7]. Эти два подхода можно использовать, последовательно чередуя извлечение знаний с этапом их формирования при решении практических задач. В результате такого чередования получается итерационная процедура по определению набора параметров внешней среды, области допустимых значений и ранга для каждого из них.

При работе с набором выявленных факторов транскрипции в рамках механизма генов-переключателей предлагается следующее разбиение на классы:

- факторы транскрипции 1-го порядка (базовые, системообразующие параметры):  $t_1^1, t_2^2, \dots, t_p^p$ ;
- факторы транскрипции 2-го порядка (основные регулирующие параметры при заданных факторах транскрипции 1-го порядка):  $t_1^1, t_2^2, \dots, t_b^b$ ;
- редкие и специальные факторы транскрипции (служат для учета редких, но важных для решения задачи условий, а также для отражения специфики различных состояний внешней среды при заданных значениях факторов транскрипции 1-го и 2-го порядков):  $t_1^1, t_2^2, \dots, t_s^s$ ;
- резервные факторы транскрипции (факторы транскрипции, неактивные при обычном режиме настройки или работы системы искусственного интеллекта):  $t_1^1, t_2^2, \dots, t_e^e$ .

$f + b + s + e = k$  — общее количество факторов транскрипции в предметной области,  $r_i$  — ранг  $i$ -го фактора транскрипции. Если стоит задача предоставления избирательного доступа к личной информации, то в качестве примера факторов транскрипции можно перечислить виды групп пользователей, которые могут иметь доступ к данным.

## 2. Определение свойств системы искусственного интеллекта (СИИ)

Основу генома СИИ составляют те свойства, которыми она должна обладать. Базовую часть генома определяет поставленная перед СИИ задача. Также на количество генов-признаков оказывают влияние заданные условия работы, они могут накладывать ограничения на количество свойств, которыми будет обладать СИИ в итоге. Задача определения базовой части генома сводится к нахождению в заданных условиях работы СИИ набора генов-признаков, кодирующих свойства СИИ, которые помогают ей решить поставленную задачу, и к определению ранга этих свойств. Каждому свойству  $x_i$  присваивается ранг  $r_i$ ;  $r_i \in \mathbb{N}$ , где  $\mathbb{N}$  — множество натуральных чисел,  $i \in [1, l]$ . Чем важнее (нужнее) свойство, тем выше его ранг. В результате мы получаем множество  $\{r_1^1, r_2^2, \dots, r_l^l\}$  — набор свойств СИИ с присвоенными рангами,  $l$  — количество генов-свойств. В качестве примера можно привести набор элементарных стратегий по проверке клиентской информации при выдаче кредита или предоставлении страховки. Каждая стратегия в данном случае — это свойство СИИ.

## 3. Создание генома СИИ

Каждое из свойств  $r_i^i$  кодируется через ген  $x_i$ , кроме того, для повышения гибкости системы к набору из  $l$  генов  $\{x_1, x_2, \dots, x_l\}$  можно прибавить  $n - l$  резервных генов. Таким образом, получается набор из  $n$  базовых генов-свойств  $\{x_1, x_2, \dots, x_l, x_{l+1}, x_{l+2}, \dots, x_n\}$ . Учет изменения параметров предметной области (изменения условий работы) осуществляется при помощи добавления в геном генов-переключателей (энхансеров). Энхансер по своей сути является функцией от множества факторов транскрипторов:

$y_i = g_i(t_1, t_2, \dots, t_j, \dots, t_k)$ , где  $y_i$  — энхансер гена-признака  $x_i$ ,  $i \in [1, n]$ , где  $n$  — количество генов-признаков в геноме.

$t_j$  —  $j$ -й фактор транскрипции  $j \in [1, k]$ , где  $k$  — количество факторов транскрипции.

Ранг фактора транскрипции здесь не указан, чтобы не перегружать запись индексами.

$g_i$  — функция, связывающая множество факторов транскрипции  $t_1, t_2, \dots, t_j, \dots, t_k$  и энхансер  $y_i$ . В общем случае  $g_i \neq g_j$ ,  $i, j \in [1, n]$ . Для резервных генов-свойств энхансеры могут быть всегда равны нулю и исключены из обучения. Теперь сформируем геном  $\{z_1, z_2, \dots, z_n\}$  из множества генов-признаков  $\{x_1, x_2, \dots, x_l, x_{l+1}, x_{l+2}, \dots, x_n\}$  и множества генов-переключателей  $\{y_1, y_2, \dots, y_n\}$  по следующему правилу:

$$z_i = y_i \cdot x_i$$

$$\vec{z} = \{z_1, z_2, \dots, z_n\} = \{y_1 \cdot x_1, y_2 \cdot x_2, \dots, y_n \cdot x_n\},$$

где  $z_i$  —  $i$ -й составной ген генома.



В качестве заданных условий работы, как правило, рассматривают требования к скорости работы, точности результата и экономичности. Разумеется, в идеале система должна работать быстро, качественно и дешево. В формализованном виде данные параметры можно представить следующим образом:

- $a \in [\alpha_1, \alpha_2]$  — точность результата,
- $v \in [v_1, v_2]$  — скорость работы,
- $c \in [c_1, c_2]$  — затраты на работу.

Для рассматриваемой задачи эти параметры являются заданными на указанных интервалах, в ходе работы СИИ их значения могут меняться.

Необходимо отметить, что количество генов-признаков и количество факторов транскрипции (параметры  $n$  и  $k$ ) являются функциями от точности, скорости и затрат:

$$\begin{aligned}n &= n(a, v, c), \\k &= k(a, v, c).\end{aligned}$$

Учитывая изложенное выше, задачу по поиску самой эффективной СИИ мы можем сформулировать как нахождение при заданных ограничениях оптимального значения функции  $F$ :

$$\begin{aligned}\text{opt } F(\bar{z}). \\ \{\bar{z}\},\end{aligned}$$

где  $F$  — функция приспособленности (Fitness). В разных задачах вид функции  $F$  может быть различным, если в качестве  $F$  используется мера расхождения результата работы СИИ со значением целевой функции, то будет осуществляться поиск минимального значения  $F$ , в других задачах, например по поиску каких-то вредоносных объектов в предметной области, будет решаться задача максимизации  $F$ .

После описанных этапов применяются классические этапы метода генетических алгоритмов (генерация начальной популяции, кроссинговер, селекция, мутации).

Изложенный подход вполне применим для построения массовых коробочных систем защиты информации частных пользователей. Благодаря таким свойствам, как самообучаемость и способность к адаптации, в сочетании с более низкой стоимостью по сравнению с другими СИИ этот класс программных приложений в сфере безопасности информации может открыть новые перспективы для развития программного обеспечения в данном направлении.

## СПИСОК ЛИТЕРАТУРЫ:

1. Солоуев Д. Конец приватности // В мире науки. 2008. № 12. С. 73–77.
2. Дайсон Э. Размышления о приватности 2.0 // В мире науки. 2008. № 12. С. 22–27.
3. Николенко С. И., Тулупьев А. Л. Самообучающиеся системы. М.: МНЦМО, 2009. — 288 с.
4. Sudarsan N., Hammond M. C., Block K. F., Welz R., Barrick J. E., Roth A., Breaker R. R. Tandem Riboswitch Architectures Exhibit Complex Gene Control Functions // Science. 2006. V. 314. P. 300–304.
5. Prud'homme B., Gompel N., Sean B. Carroll S. B. Emerging Principles of Regulatory Evolution // Proceedings of the National Academy of Sciences USA. Vol. 104. Supplement 1. March 2007. P. 206–216.
6. Гаранина М. Л. Применение механизма генов-переключателей при разработке систем искусственного интеллекта // IX Международная научно-техническая конференция «Проблемы информатики в образовании, управлении, экономике и технике». Сборник статей. С. 111–113.
7. Гаранина М. Л. Параметризация предметной области при использовании механизма генов-переключателей в ходе разработки систем искусственного интеллекта.
8. Частиков А. П., Гаврилова Т. А., Белов Д. Л. Разработка экспертных систем. Среда CLIPS. СПб.: БХВ-Петербург, 2003. — 608 с.: илл.