

Нейроинформатика, её приложения и анализ данных

XXVI Всероссийский семинар

**Красноярск
2018**

Министерство науки и высшего образования РФ
Сибирское отделение РАН
Российская ассоциация нейроинформатики
Институт вычислительного моделирования СО РАН

НЕЙРОИНФОРМАТИКА, ЕЁ ПРИЛОЖЕНИЯ И АНАЛИЗ ДАННЫХ

МАТЕРИАЛЫ
ДВАДЦАТЬ ШЕСТОГО ВСЕРОССИЙСКОГО СЕМИНАРА
28 сентября – 30 сентября 2018 года

Красноярск 2018

УДК 007.52 (042.3)

Н 43

Нейроинформатика, её приложения и анализ данных: Материалы XXVI Всероссийского семинара, 28 сентября – 30 сентября 2018 года /

Под ред. А.Н.Горбаня, отв. за вып. М.Ю. Сенашова; – Красноярск:

Институт вычислительного моделирования СО РАН, 2018. – 165 с.

В сборнике представлены материалы XXVI Всероссийского семинара «Нейроинформатика, её приложения и анализ данных», проходившей в городе Красноярске 28 сентября – 30 сентября 2018 года.

Основной задачей конференции является всесторонний и высококвалифицированный обмен новейшими достижениями в различных областях нейроинформатики (как теории нейросетей, так и в области различных практических приложений), а также в области нелинейного статистического анализа многомерных данных, обладающих нетривиальными структурами.

Большое внимание уделено анализу областей применимости и точности методов обработки многомерных данных, анализу устойчивости различных новых (нелинейных) методов кластеризации, разбор большого числа конкретных случаев, иллюстрирующих эти проблемы и достижения.

Материалы предназначены для научных работников, преподавателей, студентов и аспирантов соответствующих специальностей.

Редакционная коллегия:

Горбань Александр Николаевич – ответственный редактор

Сенашова Мария Юрьевна – ответственный за выпуск

© ИВМ СО РАН, 2018

© Коллектив авторов, 2018

ISBN 978-5-906455-14-7



9 785906 455147

НЕЙРОННЫЕ СЕТИ ВСТРЕЧНОГО РАСПРОСТРАНЕНИЯ В СИСТЕМЕ УПРАВЛЕНИЯ ЭКЗОСКЕЛЕТОМ ПО СХЕМЕ СПЕЦИАЛИЗИРОВАННОГО ОБУЧЕНИЯ

Али Яхья Сенан Аль-Баред¹, В.В. Жилин², Д.С. Кондрашов²

¹Российский Университет Дружбы Народов, кафедра кибернетики и мехатроники, SFilist@gmail.com

²Юго-Западный государственный университет, кафедра биомедицинской инженерии, kondrashov012@mail.ru

Вертикализация экзоскелета относится к многосвязным нелинейным системам управления со сложной динамикой. Математическое моделирование движения пациента в экзоскелете показало существенную зависимость траектории движения центра масс системы от параметров PI регулятора. Для получения заданных показателей качества управляемой механической системы необходимо определение оптимальных параметров регулятора, обеспечивающих минимальную колебательность устройства и минимальные отклонения от заданной траектории. Для формирования управляющих напряжений $\bar{u} = [u_1 \dots u_3]^T$ с учетом корректировки задающих воздействий используют адаптивный алгоритм формирования управляющих напряжений с учетом корректировки задающих воздействий. На рисунке 1 приведена структурная схема системы автоматического управления, реализующая этот алгоритм [1].

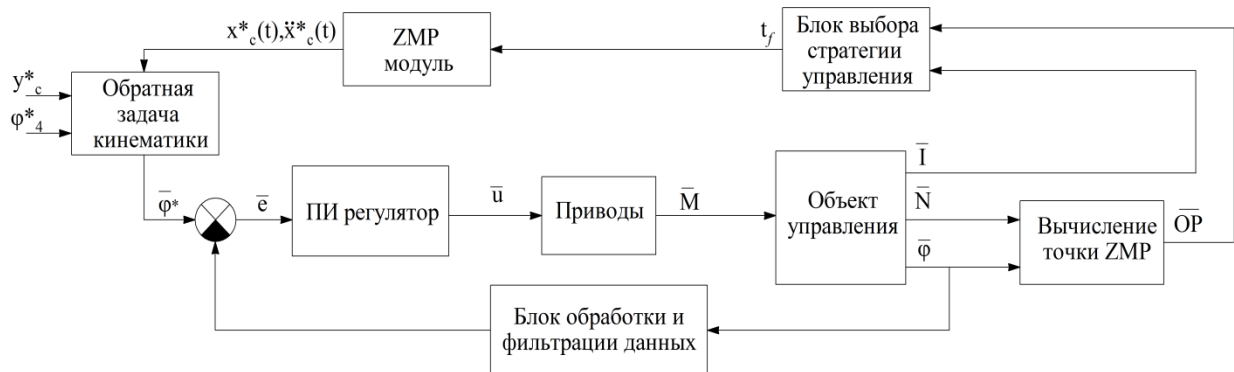


Рис. 1. Структурная схема системы автоматического управления экзоскелетом для реабилитации нижних конечностей человека

На схеме приняты следующие обозначения: $\bar{\varphi}^* = [\varphi_2^*, \varphi_3^*, \varphi_4^*]^T$ – вектор углов, задающих координаты экзоскелета в вертикальной плоскости; $\bar{\varphi} = [\varphi_2, \varphi_3, \varphi_4]^T$ – вектор углов, полученных в результате обработки сигналов с датчиков; $\bar{N} = [N_1, N_2]^T$ – вектор модулей нормальных реакций, измеренных датчиками давления, размещенными на стопах; $\bar{u} = [u_1 \dots u_3]^T$ – вектор управляющих напряжений, подаваемых на приводы; $\bar{I} = [I_1 \dots I_3]^T$ – вектор токов, протекающих в обмотках якорей электроприводов; $\bar{e} = \bar{\varphi} - \bar{\varphi}^*$ – ошибка; $\bar{M} = [M_{12}, M_{23}, M_{34}]^T$ – моменты, создаваемые электроприводами.

Для управления приводом экзоскелета использовалась последовательная схема нейронного управления [2, 3]. В качестве нейросетевого управления использовалась последовательная схема нейросетевого управления. Для реализации этой схемы

управления нейронная сеть должна обеспечить обратную динамику объекта управления. Сущность реализации обратной динамики иллюстрирует схема, представленная на рисунке 2. В этой схеме сначала настраивают NET1, а затем ее копируют в NET2 согласно схеме рисунок 2.

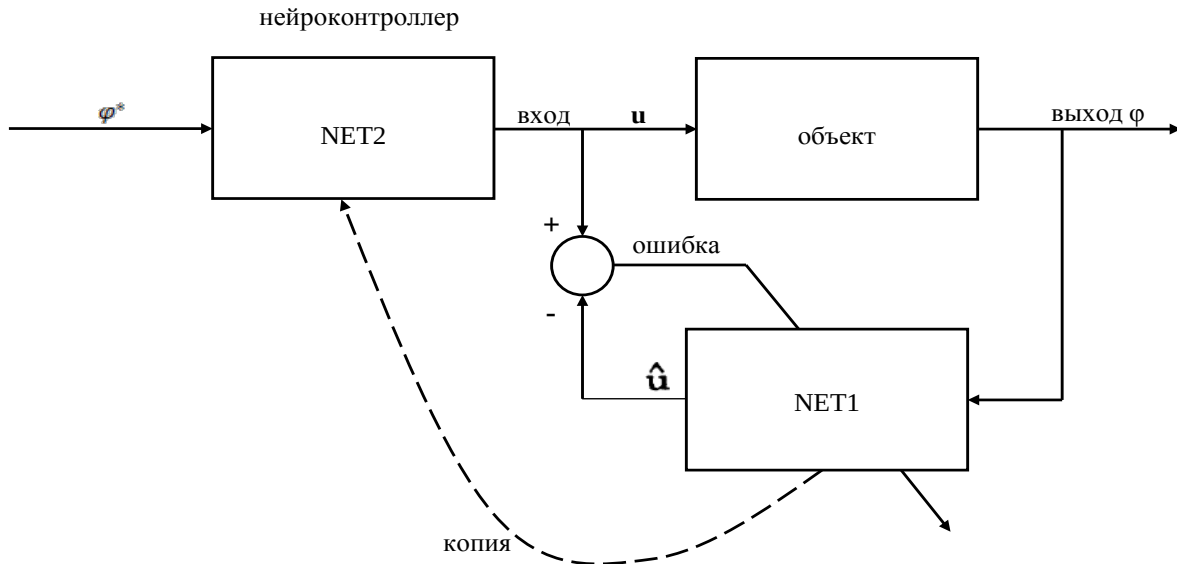


Рис. 2. Схема реализации инверсной динамики с помощью нейронных сетей

Если обозначить соотношение между входом и выходом нейронной сети NET2 как $\varphi = f_p(u)$, то нейронная сеть NET1 реализует обратное отображение $u = f_p^{-1}(\varphi^*)$, где f_p^{-1} - отображение, обратное отображению f_p . То есть, если на входе NET2 сигнал φ^* , то выходной сигнал объекта управления φ принимает значение φ^* , так как $\varphi = f_p(u) = f_p(f_p^{-1}(\varphi^*)) = \varphi^*$.

Однако в схеме рисунок 2 для обучения NET2 нельзя применить классический метод обратного распространения с аналитически заданным градиентом, так как между выходом системы управления и выходом NET2 стоит объект управления, якобиан которого в общем случае неизвестен. Поэтому приходится применять численные методы аппроксимации якобиана системы или модифицировать метод обратного распространения ошибки с учетом отсутствия информации о якобиане системы. В связи с этим было принято решение использовать в схеме специализированного обучения [2, 3]. Фрагмент этой схемы показан на рис. 3.

Если известна математическая модель объекта, то, решив соответствующую систему дифференциальных уравнений численными методами, получаем пары координат $[u(t), \varphi(t)]$, которые будем использовать для настройки нейронной сети NET1, которая в данном случае будет являться эмулятором объекта управления. Однако после настройки мы не можем использовать эту нейронную сеть в последовательной схеме нейросетевого управления, так как для этого необходима нейронная сеть, отражающая инверсную, а не прямую, как в данном случае, динамику объекта управления. Для инверсии нейронной сети NET1 используем нейронную сеть встречного распространения (ВР), которая является универсальным инвертором [4].

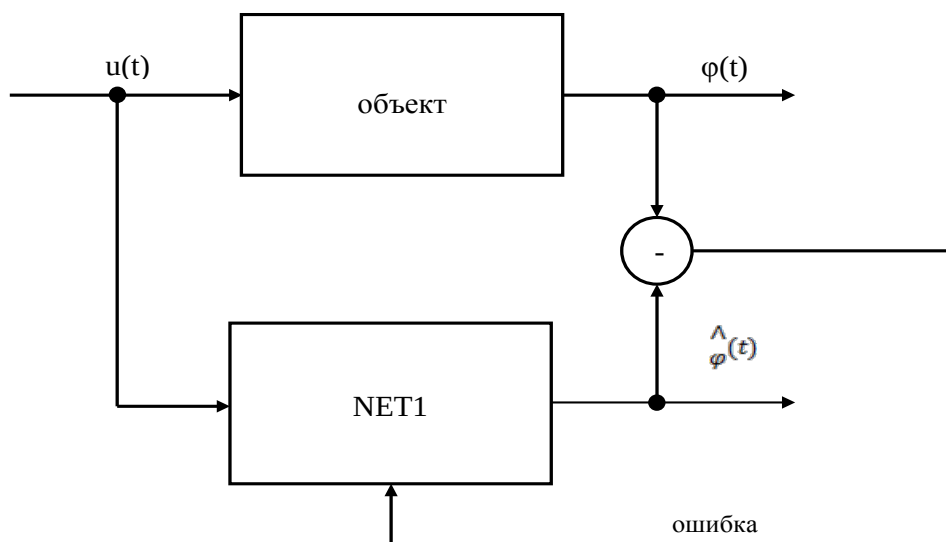


Рис. 3. Фрагмент схемы специализированного обучения

Сеть ВР состоит из двух слоёв: слоя Кохонена и слоя нейронов Гроссберга (рис. 4). На вход сети поступает пара векторов $\bar{\varphi}$ и $\bar{u}$. На выходе получаем их копии $\hat{\bar{\varphi}}$ и $\hat{\bar{u}}$.

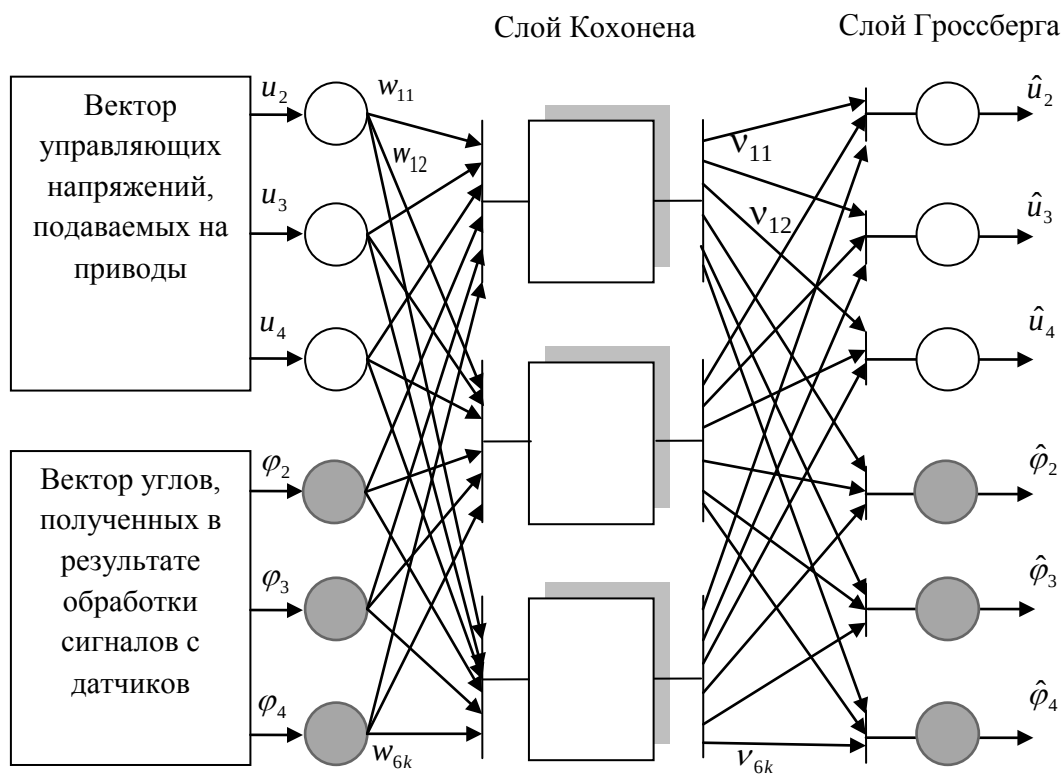


Рис. 4. Сеть встречного распространения ошибки

Если предварительно обученную ВР включить в последовательную схему нейросетевого управления, то она на выходе все равно будет отображать пару векторов $\bar{\varphi}$ и $\bar{u}$ несмотря на то, что на входе будет присутствовать только один вектор $\bar{\varphi}$. Использование сети ВР в последовательной схеме нейросетевого управления по схеме специализированного обучения обуславливается её уникальной способностью порождать функцию и обратную к ней. Следовательно, после обучения сети имеется возможность предъявлять на её вход только вектор $\bar{\varphi}$, заполняя вектор $\bar{u}$ нулевыми элементами, получая при этом на выходе вектор управляющих напряжений, на который настроена нейронная сеть. При этом размерности векторов $\bar{\varphi}$ и $\bar{u}$ могут не совпадать, что характерно для многосвязных систем управления.

Формат исходных данных для обучения сети ВР представлен в таблице. В строке этой таблицы представлен соответствующий пример, который подаётся на вход сети Кохонена при обучении сети ВР.

Таблица

Формат данных для обучения сети встречного распространения ошибки

X_1	X_2	...	X_n	Y_1	Y_2	...	Y_m
x_1^1	x_2^1	...	x_n^1	y_1^1	y_2^1	...	y_m^1
x_1^2	x_2^2	...	x_n^2	y_1^2	y_2^2	...	y_m^2
...	...	...	...	...	...	...	...
x_1^q	x_2^q	...	x_n^q	y_1^q	y_2^q	...	y_m^q
...	...	...	...	...	...	...	...
x_1^Q	x_2^Q	...	x_n^Q	y_1^Q	y_2^Q	...	y_m^Q

В результате обучения входные векторы ассоциируются с выходными и, когда сеть обучена, подача входных образов приводит к получению выходных. Правильный выход может быть получен и тогда, когда вход является неполным, несколько неверным.

Следовательно, в базе данных для обучения сети ВР должны храниться примеры, полученные в результате решения системы дифференциальных уравнений модели или полученные на основе экспериментальных исследований экзоскелета.

Подобно другим сетям сеть ВР работает в двух режимах: обучение и использование. В первом случае на входы подаём и вектор X , и вектор Y , корректируем весовые коэффициенты; во втором режиме на вход уже обученной сети подаём X или Y , а с выхода снимаем и X , и Y .

Общий алгоритм функционирования сети: на вход подаём первый образ Z .

Образ Z формируется как объединение векторов X и Y :

$$Z = X \cup Y. \quad (1)$$

В каждом нейроне k слоя Кохонена рассчитывается активация

$$A_k = \sum_{i=1}^{n+m} w_{ik} z_i = W^T Z. \quad (2)$$

Только на выходе одного нейрона j слоя Кохонена будет единица, а именно

$$OUT_j^K = \begin{cases} 1, & \text{если } A_j = \max_k A_k \\ 0, & \text{в противном случае.} \end{cases} \quad (3)$$

Очевидно, что веса, связанные с выигравшим нейроном, максимально близки к компонентам входного вектора. Следующий шаг – коррекция весовых коэффициентов, рёбра которых соответствуют выигравшему нейрону.

Слой Гроссберга функционирует подобно слою Кохонена. Активация нейрона g его слоя

$$B_g = \sum_{k=1}^K v_{kg} OUT_k^K. \quad (4)$$

На выходе нейронов слоя Гроссберга будут веса, соответствующие “выигравшему” нейрону слоя Кохонена $OUT_j^G = B_j$. Далее корректируем веса, инцидентные этому нейрону и подаём на вход сети следующий образ. В результате многократного выполнения указанной процедуры весовые коэффициенты слоя Гроссберга должны совпадать или быть достаточно близкими к входным образам.

Рассмотрим нюансы функционирования сети ВР в следующем алгоритме. Напомним, что исходные данные представлены в таблице 1 и один входной образ соответствует одной строке этой таблицы.

Алгоритм функционирования сети встречного распространения включает следующие шаги.

Шаг 1. Нормируем каждый элемент таблицы, так, например, нормированный аналог второго элемента первой строки рассчитываем по формуле

$$z_2^1 = \frac{x_2^1}{\sqrt{\sum_{j=1}^n (x_j^1)^2 + \sum_{j=1}^m (y_j^1)^2}} \quad (5)$$

и заменяем в таблице 1 каждый элемент на нормированный аналог.

Шаг 2. Генерируем случайным образом весовые коэффициенты и обязательно их нормализуем, тем самым сокращая процесс обучения.

Шаг 3. Подаём на вход сети строку матрицы Z и рассчитываем скалярные произведения с векторами весов, которые связаны со всеми нейронами Кохонена.

Шаг 4. Среди всех скалярных произведений выбираем наибольшие и подстраиваем веса соответствующего нейрона согласно выражению $W_h = W_G + \alpha(Z - W_G)$, где W_G - предыдущее значение веса, W_h - его новое значение, α - коэффициент обучения, вначале примерно равный 0,7 и постепенно уменьшающий в

процессе обучения. Выход “выигравшего” нейрона равен единице, всех остальных равен нулю.

Шаг 6. Корректируем все веса слоя Гроссберга в соответствии с выражением:

$$V_{ijh} = V_{ijG} + \beta \cdot (Z_j - V_{ijG}) \cdot \chi_i^K, \quad (6)$$

где β - коэффициент обучения (вначале принимается равным 0,1), Z_j – j -ая компонента реального вектора выхода, V_{ijG} - старое значение весового коэффициента слоя Гроссберга, V_{ijh} - новое значение,

$$\chi_i^K = \begin{cases} 1, & \text{если } i - \text{й нейрон слоя Кохонена "выиграл"}, \\ 0, & \text{в противном случае.} \end{cases} \quad (7)$$

Шаг 7. Уменьшить значения α и β .

Шаг 8. Повторять шаги 3...7 до тех пор, пока каждая входная пара из обучающей выборки не будет порождать аналогичную пару на выходе.

Таким образом, предложена новая структура последовательной схемы нейросетевого управления, построенная по схеме специализированного обучения с использованием нейронной сети встречного распространения ошибки в качестве нейроконтроллера. Схема нейросетевого управления апробирована на прототипе экзоскелета при управлении вертикализации пациента.

Список литературы

1. Яцун А.С. Адаптивная система управления экзоскелета, осуществляющая вертикализацию человека/А.С. Яцун, С.И. Савин, Р.Н. Турлапов //Известия Юго-Западного государственного университета. - 2015. - №3. – С.26-32.
2. Сигеру Омату. Нейроуправление и его приложение. Кн. 2/Сигеру Омату, Марзуки Халид, Рубия Юсоф. Пер. с англ. – М. ИМПЖ. 2000. – 272 с.
3. Терехов В.А. Нейросетевые системы управления: Учеб. пособие для вузов/В.А. Терехов, Д.В. Ефимов, И.Ю. Тюкин. М.: Высш. шк.2002.-183 с.
4. Филист С.А. Моделирование морфологических образований на рентгенограммах грудной клетки в интеллектуальных диагностических системах медицинского назначения/С.А. Филист, П.С. Кудрявцев, А.А. Кузьмин и др.// Прикаспийский журнал: управление и высокие технологии. 2017. № 3 (39). - С. 109-120.
5. Филист С.А. Формирование признакового пространства для задач классификации сложноструктурируемых изображений на основе спектральных окон и нейросетевых структур/ С.А. Филист, Кузьмин А.А., Шаталова О.В. и др.//Известия Юго-Западного государственного университета. 2016. № 4 (67). - С. 56-68.

РЕШЕНИЕ КРАЕВЫХ ЗАДАЧ НА СЕТЯХ РАДИАЛЬНЫХ БАЗИСНЫХ ФУНКЦИЙ

М.М. Алкезуини, В.И. Горбаченко

Пензенский государственный университет,
факультет вычислительной техники,

кафедра "Компьютерные технологии", gorvi@mail.ru

Для численного решения краевых задач, описываемых дифференциальными уравнениями в частных производных (ДУЧП), получают популярность бессеточные методы [1], не требующие сложной процедуры построения сетки. Среди бессеточных методов особой популярностью пользуются проекционные методы, использующие радиальные базисные функции (РБ-функции) [2–4] в качестве базисных функций. Методы, использующие РБ-функции, позволяют получить дифференцируемое решение в любой точке области решения, применимы к задачам произвольной размерности со сложными расчетными областями. При решении задачи параметры базисных функций выбираются, а веса вычисляются так, чтобы невязки, получаемые после подстановки приближенного решения в пробных точках в уравнение, были равны нулю. Главный недостаток таких методов — необходимость неформального подбора параметров базисных функций. Этот недостаток преодолевается применением специального класса нейронных сетей — сетей радиальных базисных функций (РБФ-сетей) [5–7]. Применение РБФ-сетей позволяет в процессе обучения сетей настраивать как веса, так и параметры РБ-функций.

Решение задачи формируется в процессе обучения РБФ-сети. Поэтому актуальной является разработка быстрых алгоритмов обучения РБФ-сетей при решении краевых задач. Целью данной работы является адаптация современных ускоренных градиентных методов первого порядка и метода Левенберга-Марквардта для обучения РБФ-сетей при решении краевых задач. Данная работа является развитием [8], в которой перечисленные методы применены для решения задач аппроксимации функций с помощью РБФ-сетей.

РБФ-сеть включает два слоя [9]. Первый слой состоит из РБ-функций, производящих нелинейное преобразование входного вектора $\mathbf{x} = [x_1, x_2, \dots, x_d]$ — координат точки, в которой вычисляется приближение к решению (d — размерность пространства). РБ-функция [3] — это функции расстояния точки пространства от параметра функции, называемого центром функции: $\varphi(\|\mathbf{x} - \mathbf{c}\|, \mathbf{p})$, где $\mathbf{x}$ — точка пространства, $\mathbf{c}$ — центр радиальной базисной функции, $\|\mathbf{x} - \mathbf{c}\|$ — евклидова норма (расстояние) между точкой и центром, $\mathbf{p}$ — вектор параметров функции. Применяются различные РБ-функции. В данной работе используется функция Гаусса (Гауссиан) $\varphi(\|\mathbf{x} - \mathbf{c}\|, a) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{2a^2}\right)$, где $\mathbf{c}$ — положение центра функции, a — параметр формы, часто называемый шириной. Второй слой РБФ-сети представляет собой линейный взвешенный сумматор, при решении краевых задач описываемый выражением

$$u(\mathbf{x}) = \sum_{k=1}^M w_k \varphi_k(\mathbf{x}, \mathbf{p}_k), \quad (1)$$

где M — количество РБ-функций, w_k — вес РБ-функции φ_k , $\mathbf{p}_k$ — вектор параметров РБ-функции φ_k .

Решение краевых задач на РБФ-сетях рассмотрим на примере решения стационарной краевой задачи, описывающей процесс в однородной среде

$$Lu(\mathbf{x}) = f(\mathbf{x}), \mathbf{x} \in \Omega; Bu(\mathbf{x}) = p(\mathbf{x}), \mathbf{x} \in \partial\Omega, \quad (2)$$

где u — искомое решение; L — дифференциальный оператор, например, $L = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$; B — оператор граничных условий; Ω — область решения; $\partial\Omega$ — граница области; f и p — известные функции.

Обучение сети, т.е. настройка весов $\mathbf{w}$ и параметров $\mathbf{p}$ РБ-функций, основано на минимизации функционала ошибки, представляющего собой сумму квадратов невязок в пробных точках для уравнения и граничных условий

$$I(\mathbf{w}, \mathbf{p}) = \sum_{i=1}^N [Lu_{\text{RBF}}(\mathbf{x}_i; \mathbf{w}, \mathbf{p}) - f(\mathbf{x}_i)]^2 + \lambda \sum_{i=N+1}^{N+K} [Bu_{\text{RBF}}(\mathbf{x}_i; \mathbf{w}, \mathbf{p}) - p(\mathbf{x}_i)]^2 \rightarrow \min, \quad (3)$$

где $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N \in \Omega$, $\mathbf{x}_{N+1}, \mathbf{x}_{N+2}, \dots, \mathbf{x}_{N+K} \in \partial\Omega$; λ — подбираемый штрафной множитель, учитывающий вклад невязок в граничных пробных точках, u_{RBF} — приближение к решению (1), полученное РБФ-сетью, N, K — количество пробных точек внутри и на границе области решения соответственно.

Для обучения РБФ-сетей применяются градиентные методы. Среди градиентных методов выделяют методы нулевого порядка, использующие только информацию о значении минимизируемой функции, методы первого порядка, использующие первые производные минимизируемой функции, т.е. градиент функции, и методы второго порядка, использующие вторые производные — матрицу Гессе. В известных работах [5] используется метод первого порядка — метод градиентного спуска. Для описания алгоритма обучения сети рассмотрим вектор $\boldsymbol{\theta}$ параметров сети, включающий веса сети и параметры РБ-функций. Тогда настройка вектора $\boldsymbol{\theta}$ параметров РБФ-сети на k -й итерации градиентного спуска описывается следующим образом

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \Delta\boldsymbol{\theta}^{(k+1)}, \quad (4)$$

где $\Delta\boldsymbol{\theta}^{(k+1)} = -\eta \nabla I(\boldsymbol{\theta}^{(k)})$ — поправка вектора $\boldsymbol{\theta}$, η — подбираемый числовой коэффициент (скорость обучения), $\nabla I(\boldsymbol{\theta}^{(k)})$ — вектор градиента функционала (3) по значению параметра $\boldsymbol{\theta}^{(k)}$ на итерации k .

Недостатком метода градиентного спуска является низкая скорость сходимости, что не позволяет получить высокую точность решения.

Методы второго порядка не получили распространения при обучении РБФ-сетей. Метод Левенберга-Марквардта — наиболее быстрый метод второго порядка, практически не применяется для обучения РБФ-сетей. Известны только отдельные примеры применения этого метода [10–12] в областях, не связанных с решением краевых задач.

Для обучения РБФ-сетей перспективным является метод доверительных областей [13], обладающий высокой скоростью сходимости. Использование метода доверительных областей для обучения РБФ-сетей при решении краевых задач

предложено в [14] и развито в [7, 15]. Но этот метод достаточно сложен, так как на каждой итерации обучения сети приходится решать задачу условной минимизации.

Рассмотрим применение современных методов первого порядка и метода Левенберга-Марквардта для обучения РБФ-сетей при решении краевых задач. Метод градиентного спуска с импульсом [16] отличается от метода градиентного спуска другим вычислением вектора поправки в формуле коррекции вектора параметров сети (4):

$$\Delta\theta^{(k+1)} = \alpha\Delta\theta^{(k)} - \eta\nabla I(\theta^{(k)}), \quad (5)$$

где $\nabla I(\theta^{(k)})$ — градиент функционала (3), η — скорость обучения, α — коэффициент момента, принимающий значения в интервале $[0, 1]$.

Слагаемое в (5), не зависящее от градиента, оказывает наибольшее влияние на изменение вектора поправки в областях с малым значением градиента. Можно сказать, метод позволяет "по инерции" преодолевать такие области на плоских участках функционала ошибки и вблизи локальных максимумов.

При обучении нейронных сетей глубокой архитектуры популярен метод ускоренного градиента Нестерова (NAG — Nesterov Accelerated Gradient) [16–17]. В методе NAG поправка вектора параметров сети описывается следующим образом

$$\Delta\theta^{(k+1)} = \alpha\Delta\theta^{(k)} - \eta\nabla I(\theta^{(k)} + \alpha\Delta\theta^{(k)}).$$

Метод NAG отличается от градиентного спуска с импульсом более точным вычислением вектора градиента: градиент вычисляется в точке, учитывающей коррекцию параметров.

В реализации методов первого порядка компоненты вектора градиента функционала ошибки по весам, центрам и ширине несложно вычислить аналитически. В качестве примера рассмотрим модельную задачу, описываемую уравнением Лапласа с граничными условием Дирихле

$$\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = f(x_1, x_2), (x_1, x_2) \in \Omega, u = p(x_1, x_2) \in \partial\Omega, \quad (6)$$

где $\partial\Omega$ — граница области; f и p — известные функции (x_1, x_2) .

Тогда функционал ошибки (3) запишется в виде

$$I = \sum_{i=1}^N (\Delta u_i - f_i)^2 + \lambda \sum_{j=1}^K (u_j - p_j)^2, \quad (7)$$

где Δu_i — значение Лапласиана в точке i .

Компоненты градиента по весам РБФ-сети вычисляются по формуле

$$\frac{\partial I}{\partial w_p} = \sum_{i=1}^N (\Delta u_i - f_i) e^{-\frac{\|x_i - c_p\|^2}{2a_p^2}} \frac{\|x_i - c_p\|^2 - 2a_p^2}{a_p^4} + \lambda \sum_{j=1}^K (u_j - p_j) e^{-\frac{\|x_j - c_p\|^2}{2a_p^2}}.$$

Компонент градиента по координате c_{p1} центра РБ-функций имеет вид

$$\frac{\partial I}{\partial c_{p1}} = w_p \sum_{i=1}^N (\Delta u_i - f_i) e^{-\frac{\|x_i - c_p\|^2}{2a_p^2}} (x_{i1} - c_{p1}) \frac{\|x_i - c_p\|^2 - 4a_p^2}{a_p^6} +$$

$$\lambda w_p \sum_{j=1}^K (u_j - p_j) e^{-\frac{\|x_j - c_p\|^2}{2a_p^2}} \frac{(x_{j1} - c_{p1})}{a_p^2}.$$

Аналогичный вид имеет компонент градиента по координате c_{p2} .

Компонент градиента по ширине равен

$$\frac{\partial I}{\partial a_p} = w_p \sum_{i=1}^N (\Delta u_i - f_i) e^{-\frac{\|x_i - c_p\|^2}{2a_p^2}} \frac{\|x_i - c_p\|^4 - 6a_p^2 \|x_i - c_p\|^2 + 4a_p^4}{a_p^7} +$$

$$\lambda w_p \sum_{j=1}^K (u_j - p_j) e^{-\frac{\|x_j - c_p\|^2}{2a_p^2}} \frac{\|x_j - c_p\|^2}{a_p^3}.$$

Процесс обучения сети заканчивается при малом значении функционала ошибки (7) или средней квадратической погрешности.

Рассмотрим применение для обучения сети метода второго порядка — метода Левенберга-Марквардта. Рассмотрим задачу (6). Введем вектор параметров сети

$$\boldsymbol{\theta} = [w_1, w_2, \dots, w_M, c_{11}, c_{21}, \dots, c_{M1}, c_{12}, c_{22}, \dots, c_{M2}, a_1, a_2, \dots, a_M]^T,$$

где параметры j -й РБ-функции ($j = 1, 2, 3, \dots, M$): w_j — вес, c_{j1} и c_{j2} — координаты центра, a_j — ширина.

В методе Левенберга-Марквардта поправка $\Delta \boldsymbol{\theta}^{(k+1)}$ вектора параметров $\boldsymbol{\theta}$ находится из решения системы линейных алгебраических уравнений

$$(\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{E}) \Delta \boldsymbol{\theta}^{(k+1)} = -\mathbf{g}_k, \quad (8)$$

где матрица $\mathbf{J}_k^T \mathbf{J}_k + \mu_k \mathbf{E}$ является аппроксимацией матрицы Гессе, $\mathbf{E}$ — единичная матрица, μ_k — параметр регуляризации, изменяющийся на каждом шаге обучения, $\mathbf{g} = \mathbf{J}^T \mathbf{e}$ — вектор градиента функционала (7) по вектору параметров $\boldsymbol{\theta}$, $\mathbf{e} = [e_1, e_2, \dots, e_{N+K}]^T$ — вектор невязок во внутренних и граничных пробных точках, $\mathbf{J}_k$ — матрица Якоби, вычисленная по значениям параметров сети на k -й итерации.

Представим матрицу Якоби в блочном виде $\mathbf{J} = [\mathbf{J}_w \mathbf{J}_{c_1} \mathbf{J}_{c_2} \mathbf{J}_a]$. Элементы матрицы Якоби несложно вычислить аналитически. Элементы матрицы $\mathbf{J}_w$ для внутренних пробных точек вычисляются по формуле

$$\frac{\partial e_i}{\partial w_p} = e^{-\frac{\|x_i - c_p\|^2}{2a_p^2}} \frac{\|x_i - c_p\|^2 - 2a_p^2}{a_p^4}.$$

Для граничных пробных точек вычисления производятся по формуле

$$\frac{\partial e_j}{\partial w_p} = \exp\left(-\frac{\|x_j - c_p\|^2}{2a_p^2}\right).$$

Элементы матрицы $\mathbf{J}_{c_1}$ для внутренних пробных имеют вид

$$\frac{\partial e_i}{\partial c_{p1}} = w_p e^{-\frac{\|\mathbf{x}_i - \mathbf{c}_p\|^2}{2a_p^2}} (x_{i1} - c_{p1}) \frac{\|\mathbf{x}_i - \mathbf{c}_p\|^2 - 4a_p^2}{a_p^6}.$$

Для граничных точек элементы матрицы $\mathbf{J}_{c_1}$ записываются в виде

$$\frac{\partial e_j}{\partial c_{p1}} = w_p e^{-\frac{\|\mathbf{x}_j - \mathbf{c}_p\|^2}{2a_p^2}} \frac{(x_{j1} - c_{p1})}{a_p^2}.$$

Аналогично вычисляются элементы матрицы $\mathbf{J}_{c_1}$.

Элементы матрицы $\mathbf{J}_a$ для внутренних пробных точек вычисляются по формуле

$$\frac{\partial e_i}{\partial a_p} = \frac{w_p}{a_p^5} e^{-\frac{\|\mathbf{x}_i - \mathbf{c}_p\|^2}{2a_p^2}} \left[\frac{\|\mathbf{x}_i - \mathbf{c}_p\|^2}{a_p^2} (\|\mathbf{x}_i - \mathbf{c}_p\|^2 - 2a_p^2) - 4 (\|\mathbf{x}_i - \mathbf{c}_p\|^2 - 4a_p^2) \right].$$

Для граничных точек элементы матрицы записываются в виде

$$\frac{\partial e_j}{\partial a_p} = \frac{w_p}{a_p^3} e^{-\frac{\|\mathbf{x}_j - \mathbf{c}_p\|^2}{2a_p^2}} \|\mathbf{x}_i - \mathbf{c}_p\|^4.$$

Условия завершения процесса обучения по методу Левенберга-Марквардта те же, что и в методе Нестерова.

В методе Левенберга-Марквардта параметр регуляризации μ должен изменяться в процессе обучения сети. Процесс обучения начинается при относительно большом значении параметра μ . Это означает, что в начале процесса обучения Гессиян в (8) близок к приближенному значению $\mathbf{H} \approx \mu \mathbf{E}$, а вектор поправки определяется методом градиентного спуска с малым шагом: $\Delta \boldsymbol{\theta}^{(k+1)} = -\mathbf{g}_k / \mu_k$. По мере уменьшения функционала ошибки параметр μ уменьшается и метод приближается к методу Ньютона с аппроксимацией Гессияна $\mathbf{H} \approx \mathbf{J}_k^T \mathbf{J}_k$. Это обеспечивает высокую скорость сходимости, так как метод Ньютона вблизи минимума функционала ошибки имеет хорошую сходимость.

В [18] показано, что метод Левенберга-Марквардта эквивалентен методу доверительных областей, а радиус доверительной области регулируется параметром μ . Но в отличие от известных реализаций метода доверительных областей метод Левенберга-Марквардта не требует решения на каждой итерации обучения достаточно сложной задачи условной оптимизации. То есть метод Левенберга-Марквардта сохраняя положительные свойства метода доверительных областей, является более простым.

Экспериментальное исследование рассмотренных методов проводилось на примере решения модельной задачи (6) при следующих функциях правой части и граничных условий $f(x_1, x_1) = \sin(\pi x_1) \sin(\pi x_1)$, $p(x_1, x_1) = 0$. Задача решалась в единичном квадрате. Количество внутренних пробных точек равно $N = 100$. Количество граничных пробных точек равно $K = 40$. Штрафной коэффициент функционала ошибки

по границе равен $\lambda = 10$. Центры РБ-функций располагались регулярно на квадратной сетке с количеством центров по каждой координате, равным 8. Как показали эксперименты, целесообразно расположить центры и на расстоянии одного шага сетки за пределами области решения. Пробные точки располагались случайным образом в области решения и на границе области. Веса иницировались нулевыми значениями. Начальная ширина всех РБ-функций была постоянной, равной 0,2 для всех методов. Итерационный процесс обучения продолжался до достижения заданного значения средней квадратической ошибки. Зависимость среднеквадратической ошибки различных методов от номера итерации показана на рисунке.

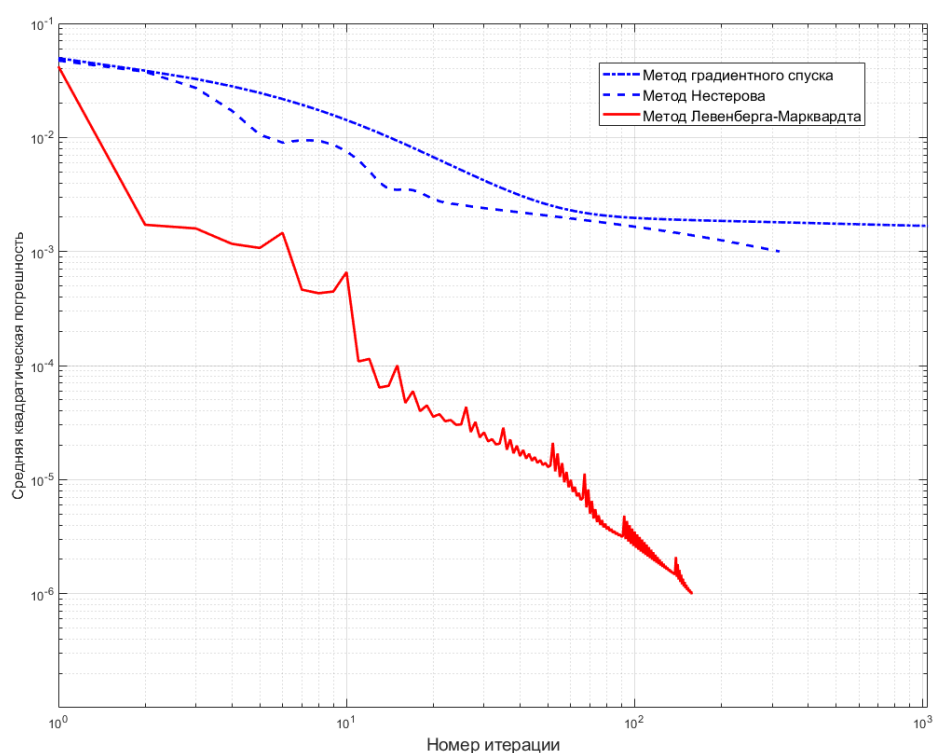


Рис. Зависимость среднеквадратической ошибки от номера итерации

Эксперименты показали качественное преимущество метода Левенберга-Марквардта перед методами первого порядка. Метод градиентного спуска позволил решить модельную задачу с небольшой точностью. Для решения с большой точностью метод практически неприменим. Несколько бóльшую точность обеспечивает метод Нестерова. Только метод Левенберга-Марквардта позволил решить задачу с высокой точностью за приемлемое время. Метод Левенберга-Марквардта показал соизмеримые результаты по сравнению с методом доверительных областей [7], но реализация метода Левенберга-Марквардта проще.

Список литературы

1. Meshfree Methods for Partial Differential Equations / Editors M. Griebel, Marc. A. Schweitzer. — Springer, 2008. — 412 p.

2. Kansa E. J. Motivation for using Radial Basis Function to solve PDEs // [Электронный ресурс]. — Режим доступа: <http://www.cityu.edu.hk/rbf-pde/files/overview-html.html>
3. Buhmann M. D. Radial Basis Functions: Theory and Implementations. — Cambridge University Press, 2004. — 259 p.
4. Chen W., Fu Z.-J. Recent Advances in Radial Basis Function Collocation Methods. — Springer, 2014. — 90 p.
5. Yadav N., Yadav A., Kumar M. An Introduction to Neural Network Methods for Differential Equations. Springer, 2015. — 115 p.
6. Васильев А. Н., Тархов Д. А., Шемякина Т. А. Нейросетевой подход к задачам математической физики. — СПб.: Нестор-История, 2015. — 260 с.
7. Горбаченко В.И., Жуков М.В. Решение краевых задач математической физики с помощью сетей радиальных базисных функций // Журнал вычислительной математики и математической физики. — 2017, Том 57. — № 1. — С.115–126.
8. Алкезуини М. М., Горбаченко В. И. Совершенствование алгоритмов обучения сетей радиальных базисных функций для решения задач аппроксимации // Модели, системы, сети в экономике, технике, природе и обществе. — 2017. — № 3 (23). — С.123–138.
9. Хайкин С. Нейронные сети: полный курс. — М.: Вильямс, 2006. — 1104 с.
10. Zhang L., Li K., He H., Irwin G.W. A New Discrete-Continuous Algorithm for Radial Basis Function Networks Construction // IEEE Trans. Neural Networks and Learning Systems. — 2013, Vol. 24. — No 11. — P. 1785–1798.
11. Xie T., Yu H., Hewlett J., Rozycki P., Wilamowski B. Fast and Efficient Second-Order Method for Training Radial Basis Function Networks // IEEE Trans. Neural Networks and Learning Systems. — 2012, Vol. 23. — No. 4. — P. 609–619.
12. An improved Levenberg-Marquardt algorithm with adaptive learning rate for RBF neural network / R. An, W. J. Li, H. G. Han, J. F. Qiao // Proceedings 2016 of the 35th Chinese Control Conference. — 3630–3635.
13. Conn A.R., Gould N.I.M., Toint P.L. Trust-Region Methods. — MPS-SIAM, 1987. — 979 p.
14. Горбаченко В. И., Жуков М. В. Обучение сетей радиальных базисных функций методом доверительных областей для решения уравнения Пуассона // Информационные технологии, 2013, №9, с. 65–70.
15. Algezweeni M. M., Gorbachenko V. I., Zhukov M. V., Jaafar M. S. Efficient Solving of Boundary Value Problems Using Radial Basis Function Networks Learned by Trust Region Method // International Journal of Mathematics and Mathematical Sciences. — Vol. 2018, Article ID 9457578, 4 pages, 2018.
16. Гудфеллоу Я., Бенджио И., Курвилль А. Глубокое обучение. — М.: ДМК Пресс, 2018.— 652 с.
17. On the importance of initialization and momentum in deep learning / I. Sutskever, J. Martens, G. Dahl, G. Hinton // Proceedings of the 30th International Conference on International Conference on Machine Learning Volume. — 2013. — Vol. 28. — P. 1139–1147.
18. Marquardt D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters // Journal of the Society for Industrial and Applied Mathematics. — 1963, Vol. 11. — No 2. — P. 431–441.

СНИЖЕНИЕ НАПРЯЖЕНИЯ ФУНКЦИОНАЛЬНЫХ СИСТЕМ ОРГАНИЗМА С ПОМОЩЬЮ САМОРЕГУЛЯЦИИ

Н.М.Богатов¹, А.Н.Богатова², С.А.Сухих³, В.В.Филимонов⁴

¹Кубанский государственный университет, bogatov@phys.kubsu.ru

²Кубанский государственный университет, anna_bogatova@mail.ru

³Кубанский государственный университет, suhih_52@mail.ru

⁴Кубанский государственный университет, vadik2393@gmail.com

Организм человека, являясь открытой энергетической системой, вступает во взаимодействие с внешней средой. Агрессивное действие внешней среды, требующее адаптации к современным социально-экономическим условиям и техногенным факторам, разрушающим экологию и биосферу, проявляется в возрастании психофизиологических, информационных нагрузок на организм. Вследствие этого повышаются биоэнергетические затраты организма, приводящее к истощению его ресурсов, появлению хронической усталости, хронического стресса, различных психологических и соматических расстройств. Для улучшения «рабочего» состояния и поддержания психического здоровья у людей, деятельность которых протекает в напряженных условиях, создаются специализированные программы и психотехнические процедуры [1].

Организм является целостной системой [2]. Саморегуляция, реализующаяся как на уровне отдельных функциональных систем организма, так и посредством их взаимодействия [3], оказывается важнейшим фактором достижения физиологической, психологической и социальной адаптации. В этом процессе важнейшую роль играют нейрогуморальные реакции [4]. Способность к саморегуляции посредством управления психоэнергетическим потенциалом является индивидуальной характеристикой человека. Положительное влияние психотехник на саморегуляцию функциональных состояний человека продемонстрировано в работах [5 – 7].

Студенты – это особая социальная группа, определяющая в перспективе трудовой и оборонительный потенциал страны, воспроизводство и здоровье будущих поколений. Условия труда и жизни студентов характеризуются повышенными психоэмоциональной, умственной нагрузками, а также необходимостью адаптации к изменяющимся условиям обучения и проживания. Сохранение здоровья студентов является составной частью общегосударственной проблемы сохранения и укрепления здоровья населения РФ.

Цель исследования – изучение методом кардиоинтервалографии влияния техник саморегуляции на функциональное состояние здоровья студентов.

В процессе эксперимента достигалось воздействие сознания субъекта на энергоинформационный обмен в организме. Информационное воздействие происходило во внутреннем пространстве субъекта и осуществлялось в результате методически организованного сеанса саморегуляции, в течение которого применялись различные психотехники. В эксперименте участвовали 22 респондента – студенты в возрасте 20 – 22 года. Участники эксперимента не обладали специальной подготовкой по управлению сознанием.

Функциональное состояние организма испытуемых оценивалось методами функциональной диагностики, основанными на измерении электрических потенциалов с помощью программно-аппаратного комплекса SilverStep. Комплекс SilverStep, используя анализ вариабельности сердечного ритма, оценивает состояние регуляторных

систем организма, в частности комплексный показатель здоровья, адаптацию, вегетативное равновесие нервной системы, спектрограмму ритмов головного мозга, нейрогуморальную регуляцию, психоэмоциональное состояние, энергетический баланс и др.

Далее использованы следующие показатели диагностического комплекса SilverStep AURA: 1 – биологический возраст; 2 – частота сердечных сокращений; 3 – мощность ауры; 4 – жизненная сила, иммунитет; 5 – мочеполовая система, почки, обмен веществ; 6 – желудочно-кишечный тракт, печень, желчный пузырь; 7 – сердечно-сосудистая и дыхательная система; 8 – эндокринная и лимфатическая система; 9 – центральная нервная система; 10 – головной мозг; 11 – Р (меридиан лёгких); 12 – GI (меридиан толстой кишки); 13 – Е (меридиан желудка); 14 – RP (меридиан селезёнки и поджелудочной железы); 15 – С (меридиан сердца); 16 – IG (меридиан тонкого кишечника); 17 – V (меридиан мочевого пузыря); 18 – R (меридиан почек); 19 – MC (меридиан головного мозга); 20 – TR (меридиан спинного мозга); 21 – VB (меридиан желчного пузыря); 22 – F (меридиан печени). Показатели диагностического комплекса SilverStep PULSE: 1 – биологический возраст; 2 – частота сердечных сокращений; 3 – А (адаптация); 4 – В (вегетативная регуляция); 5 – С (нейрогуморальная регуляция); 6 – D (психоэмоциональное состояние); 7 – Н (комплексный показатель здоровья); 8 – ИН (индекс напряжённости); 9 – ИВР (индекс вегетативного равновесия); 10 – ПНР (показатель нейрогуморальной регуляции); 11 – РО (ресурсы организма); 12 – ППС (показатель психоэмоционального состояния); 13 – СГБ (степень гармонизации биоритмов).

На Рис.1 продемонстрировано влияние психотехники упражнение «кокон» (медитация на 7 чакрах) на физиологические показатели, определяемые с помощью SilverStep AURA. Физиологические показатели испытуемых фиксировались в обычном состоянии в конце рабочего дня до сеанса саморегуляции и после.

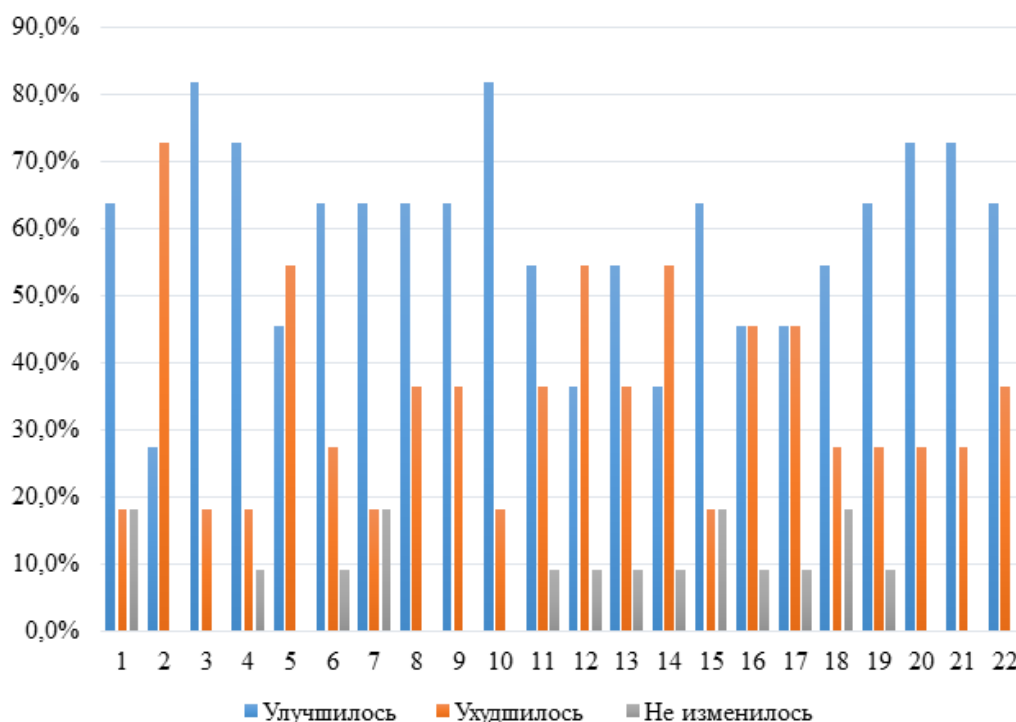


Рис.1. Изменения физиологических показателей (SilverStep AURA) в результате сеанса саморегуляции.

Гистограмма на Рис.1 показывает долю испытуемых, у которых показатели, характеризующие состояние органов и систем, пронумерованные от 1 до 22, улучшились, ухудшились или не изменились. У 65 % испытуемых улучшились показатели 6, 7, 8, 9, 15, 19, 22. Более, чем у 70 % улучшились показатели 3, 4, 10, 20, 21. Ухудшение показателей 5, 12, 14 зафиксировано у 55 %.

На Рис.2 показана гистограмма, демонстрирующая влияние психотехники упражнение «кокон» на физиологические показатели и психоэмоциональное состояние, определяемые с помощью SilverStep PULSE. Произошло явное улучшение показателей 1, 3, 4, 5, 6, 7, 10, 11, 12, 13. Показатели 2, 8, 9 отражают остаточное внутреннее напряжение после упражнения «кокон».

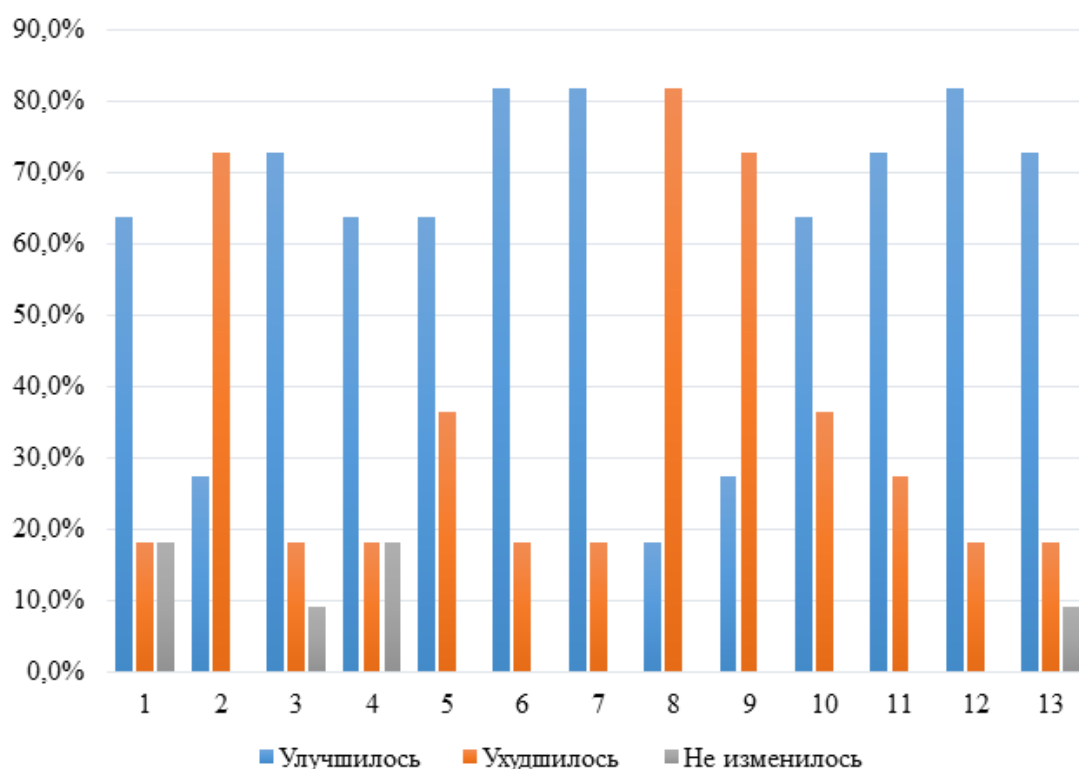


Рис.2. Изменения физиологических показателей и психоэмоционального состояния (SilverStep PULSE) в результате сеанса саморегуляции.

Повышенная умственная и нервно-психологическая нагрузка студентов приводит к истощению ресурсов организма. Исследования показывают, что многие люди реагируют на умственную деятельность в условиях дефицита времени увеличением частоты сердечного ритма и повышением артериального давления [8].

Параметры сердечного ритма используются для оценки напряжения функциональных систем организма. Разработка методик саморегуляции, направленных на сохранение функционального состояния здоровья студентов, является актуальной.

Изменение физиологических показателей и психоэмоционального состояния студентов в результате умеренной интеллектуальной нагрузки продемонстрировано на

Рис.3 и Рис.4. Испытуемые получали задание найти ответ на вопрос непосредственно в тексте учебника, не обращаясь к своим знаниям.

Из гистограммы на Рис.3 следует, что у 63,6 % студентов после интеллектуальной нагрузки участился пульс, упали показатели эндокринной и лимфатической (8), центральной нервной (9) систем, меридиана легких (11). Более, чем у 54 % испытуемых улучшились показатели мощности ауры (3), сердечно-сосудистой и дыхательной систем (7), меридиана спинного мозга (20). Таким образом, решение задач, требующих интеллектуального напряжения, оказывает явное влияние на систему кровообращения.

Из гистограммы на Рис.4 следует, что у 63,6 % студентов после интеллектуальной нагрузки участился пульс (2), ухудшились показатели нейрогуморальной (5), (10) регуляции и индекс вегетативного равновесия (9), у 68,2 % возрос индекс напряжённости. Вегетативная регуляция (4), комплексный показатель здоровья (7) стали выше более, чем у 50 % студентов. Психоэмоциональное состояние (6), (12) улучшилось у 50 %. Последние данные могут свидетельствовать о физиологической приспособленности некоторых студентов к умственным нагрузкам. У 50 % испытуемых показатель ресурсов организма (11) ухудшился после интеллектуальной нагрузки, что говорит об усталости или негативном воздействии умственной работы.

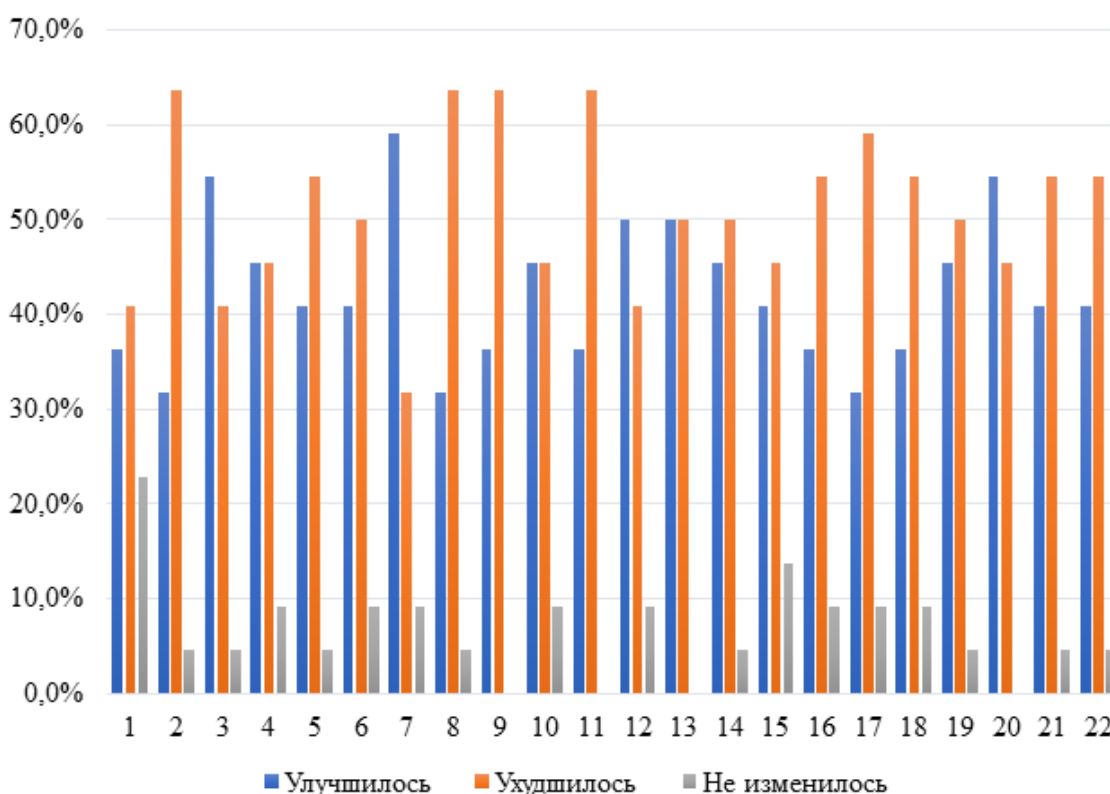


Рис.3. Изменения физиологических показателей (SilverStep AURA) в результате интеллектуальной нагрузки.

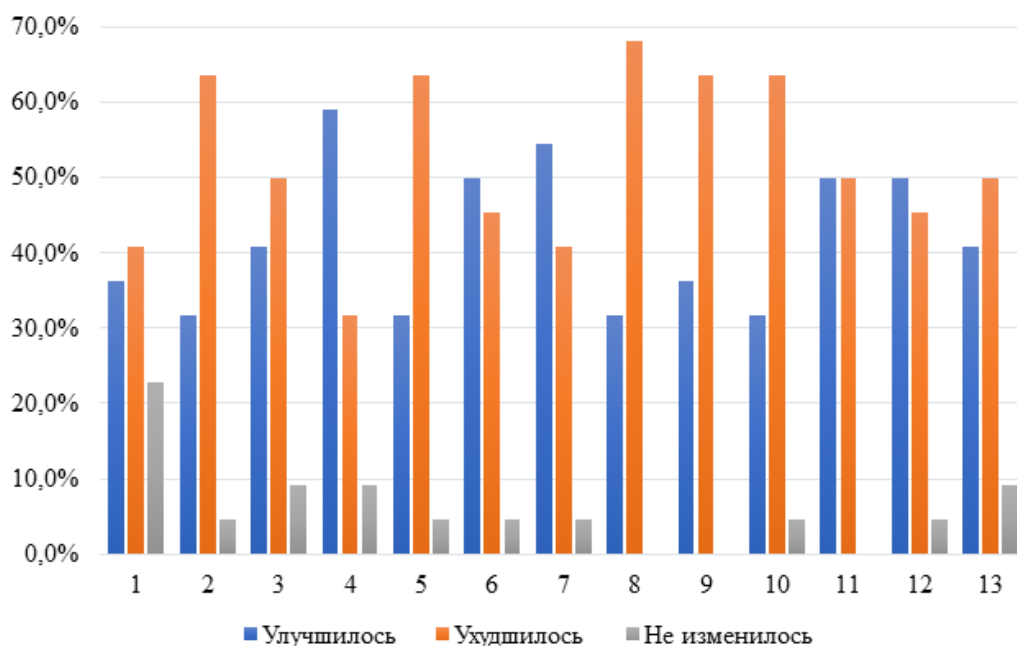


Рис.4. Изменения физиологических показателей и психоэмоционального состояния (SilverStep PULSE) в результате интеллектуальной нагрузки.

Для восстановления физиологического и психоэмоционального состояния студентов после интеллектуальной нагрузки проводился сеанс саморегуляции с использованием упражнения «кокон». Изменение физиологических показателей и психоэмоционального состояния студентов в результате восстановления продемонстрировано на Рис.5 и Рис.6.

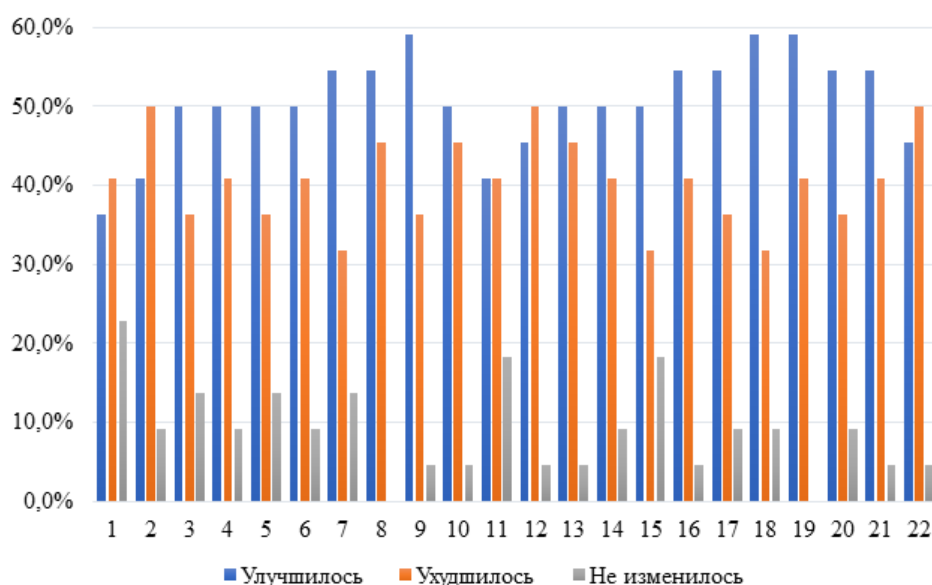


Рис.5. Изменения физиологических показателей (SilverStep AURA) в результате восстановления после интеллектуальной нагрузки.

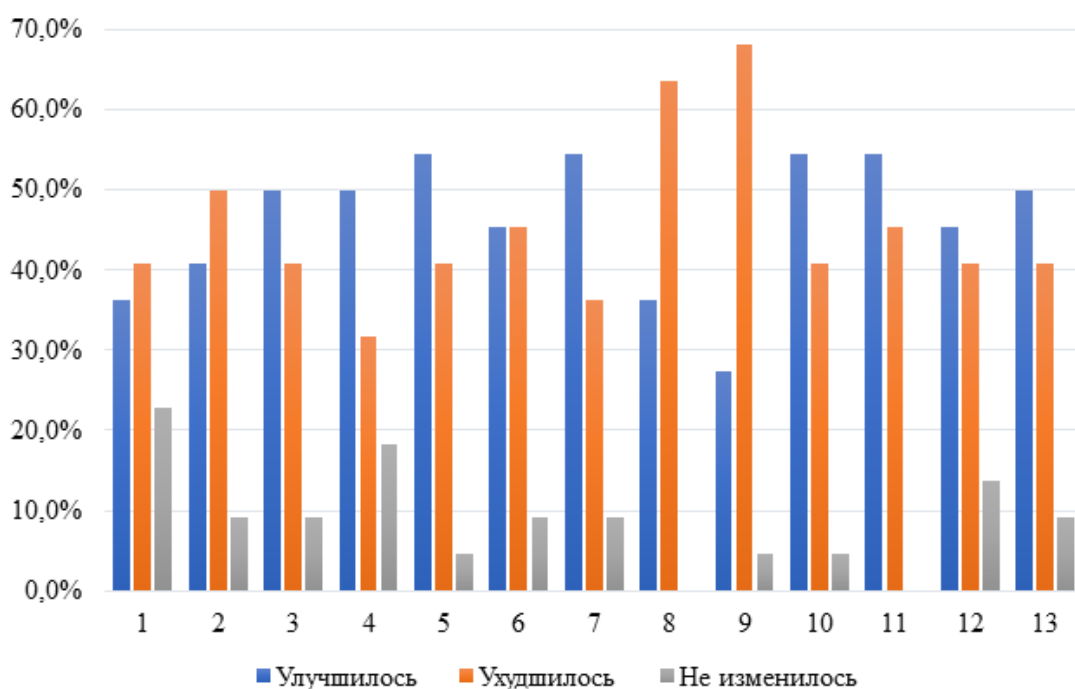


Рис.6. Изменения физиологических показателей и психоэмоционального состояния (SilverStep PULSE) в результате восстановления после интеллектуальной нагрузки.

Из гистограммы на Рис.5 следует, что у большинства студентов улучшились следующие физиологические показатели органов и систем: мощность ауры (3); жизненная сила и иммунитет (4); мочеполовая система, почки, обмен веществ (5); желудочно-кишечный тракт, печень, желчный пузырь (6); сердечно-сосудистая и дыхательная система (7); эндокринная и лимфатическая система (8); центральная нервная система (9); головной мозг (10); меридиан желудка (13); меридиан селезёнки и поджелудочной железы (14); меридиан сердца (15); меридиан тонкого кишечника (16); меридиан мочевого пузыря (17); меридиан почек (18); меридиан головного мозга (19); меридиан спинного мозга (20); меридиан желчного пузыря (21).

Из гистограммы на Рис.6 следует, что у большинства студентов улучшились следующие показатели физиологического состояния: адаптация (3); вегетативная регуляция (4); нейрогуморальная регуляция (5) и (10); комплексный показатель здоровья (7); ресурсы организма (11); степень гармонизации биоритмов (13). Показатель психоэмоционального состояния (12) улучшился у 45,5 %, ухудшился у 40,9 %, не изменился у 13,6 %. Ухудшились индексы напряженности (8) и вегетативного равновесия (9) более, чем у 60 % испытуемых.

Полученные результаты в целом свидетельствуют о снижении за счет использованной техники саморегуляции психоэмоционального и функционального напряжения организма, являющегося следствием усиления физиологических процессов, происходящих в организме в ответ на интеллектуальные нагрузки. На этапе

восстановления (сеансы саморегуляции), активность физиологических процессов снижается, что обуславливает нормализацию функционального состояния организма. Отсутствие положительного результата по некоторым показателям состояния организма требует дальнейшего изучения.

Список литературы

1. Леонова А.Б. Психологические технологии управления состоянием человека / А.Б. Леонова, А.С. Кузнецова. — М.: НПФ «Смысл», 2015. — 380 с.
2. Павлов И.П. Поли. собр. соч.: В 6 т., Т. 2, кн. 2 / И.П.Павлов. — М.-Л.: Издательство АН СССР, 1951. — 592 с.
3. Анохин П.К. Очерки по физиологии функциональных систем / П.К.Анохин. — М.: Медицина, 1975. — 448 с.
4. Данилова Н.Н. Физиология высшей нервной деятельности / Н.Н.Данилова, А.Л.Крылова. — Ростов н/Д: «Феникс», 2005. — 478 с.
5. Богатов Н.М. Исследование процессов саморегуляции организма тепловизионным методом / Н.М. Богатов, А.Н. Богатова, С.А. Сухих // Нейроинформатика, её приложения и анализ данных: Материалы XXIV Всероссийского семинара, 30 сентября – 2 октября 2016 г. Красноярск. — Красноярск: Институт вычислительного моделирования СО РАН. — 2016. — С.9 – 12.
6. Богатов Н.М. Влияние техник саморегуляции на показатели здоровья студентов / Н.М. Богатов, А.Н. Богатова, В.С. Лахин, С.А. Сухих // Нейроинформатика, её приложения и анализ данных: Материалы XXV Всероссийского семинара, 29 сентября – 1 октября 2017 г. Красноярск. — Красноярск: Институт вычислительного моделирования СО РАН. — 2017. — С.27 – 30.
7. Богатов Н.М. Влияние психотехник на саморегуляцию функционального состояния человека / Н.М. Богатов, А.Н. Богатова, С.А. Сухих // Коллективная монография. Выпуск 5. Современные проблемы физики, биофизики и инфокоммуникационных технологий. — Краснодар: ЦНТИ. — 2016. — С.5–16.
8. Горбунов Н.П. Системный подход к оценке функционального состояния в процессе учебной деятельности / Н.П. Горбунов, О.Б. Кузнецова, Л.Ю. Дорофеева // Вестник Уральской медицинской академической науки. — 2006. — Т.5. — № 3. — С.55 – 56.

МЕТОД ГЛАВНЫХ КОМПОНЕНТ В ЗАДАЧЕ АНАЛИЗА ГИДРОФИЗИЧЕСКИХ ХАРАКТЕРИСТИК ОТКРЫТЫХ ВОДОЕМОВ

О.С.Володько^{1,2}, Л.А.Компаниец^{1,2}, Л.В.Гаврилова²

¹Институт вычислительного моделирования СО РАН, osv@icm.krasn.ru,

²Сибирский федеральный университет, lv.gavrilova@gmail.com

Теоретическое обоснование метода главных компонент в общем случае и для гидрофизических характеристик относится к прошлому веку и основано на работах [1-4]. Обоснование комплексного варианта метода для анализа горизонтальных скоростей можно найти в [5]

Отметим один из последних содержательных обзоров по методу применения метода главных компонент в метеорологии и океанографии [6] и обзор философского характера [7].

В литературе известны результаты анализа методом главных компонент следующих гидрофизических величин: скорости трехмерного течения жидкости, распределения температуры по глубине и на поверхности, солёности, отклонения свободной поверхности от невозмущенного положения и т.д.

Этот метод представляет собой вариант статистической обработки данных и мы проиллюстрируем его для комплексной горизонтальной скорости.

Пусть результаты измерений записываются в виде

$$w(z_m, t_k) = u(z_m, t_k) + iv(z_m, t_k),$$

где и z_m ($m = 1, \dots, N$) — переменные по времени и пространству соответственно. Тогда поле скоростей имеет вид

$$w(x_m, t_k) = \sum_{n=1}^N E_n(t_k) \varphi_n(x_m),$$

где собственные векторы корреляционной матрицы φ_n — эмпирические ортогональные моды, а $E_n(t_k)$ — соответствующие им модальные коэффициенты.

Сумма всех собственных значений равна общей энергии:

$$\frac{1}{K} \sum_{k=1}^K \sum_{m=1}^N w^*(x_m, t_k) w(x_m, t_k) = \frac{1}{K} \sum_{k=1}^K \sum_{m=1}^N E_n^*(t_k) E_n(t_k) = \sum_{n=1}^N \lambda_n,$$

и собственные значения λ_n действительные и положительные.

Поскольку разложение формальное, то каждый раз при применении этого метода возникает необходимость интерпретации полученных слагаемых. Приведем примеры применения метода для конкретных измерений.

1. В работах [8, 9] применялись действительный и комплексный варианты метода главных компонент для оценки течений в районе мыса Хаттерас, Северная Каролина.

Первая мода для горизонтального течения здесь интерпретируется как соответствующая баротропному течению.

2. В работе [10] применялся комплексный вариант метода главных компонент для анализа горизонтальных течений в районе побережья Джорджии, строился и анализировался годограф первого модального коэффициента.

3. В работе [4] рассматривается горизонтальное течение на поверхности, и первая мода может быть интерпретирована как течение в одном направлении.

4. В работе [11] исследовалась температура на поверхности на основе многолетних измерений для Северо-Западной Атлантики. Здесь первая мода положительна и отвечает за общее нагревание и остывание.

5. При анализе долговременных значений температуры южной части Тихого океана [12] первый модальный коэффициент дает линейный тренд повышения температуры.

6. Долговременные исследования трехмерной картины распределения температуры и солёности были проведены в работах [13, 14]. В работе [15] обрабатывают трехмерные данные для солёности и температуры с использованием буев-профиломеров, которые через указанные промежутки времени поднимаются к поверхности и передают на спутник данные.

Мы применили метод главных компонент для исследования скоростей и температуры в неглубоком озере Шира (максимальная глубина 25 м) в летний период 2014-2015 гг. и сравнили эти результаты с результатами для других водоемов. Так если реальное распределение скорости в озере Шира на 01.07.2015 имеет вид как на рис.1, то при обработке данных для распределения скорости по глубине в одной точке получается первая мода, которую можно трактовать как распределение по глубине скорости течения однородной жидкости (рис.2).

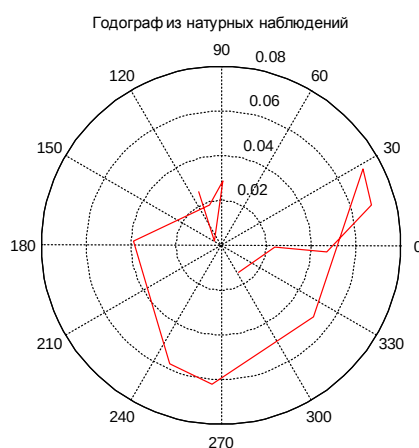


Рис.1. Распределение скорости по глубине в реальных измерениях для озера Шира (м/с).

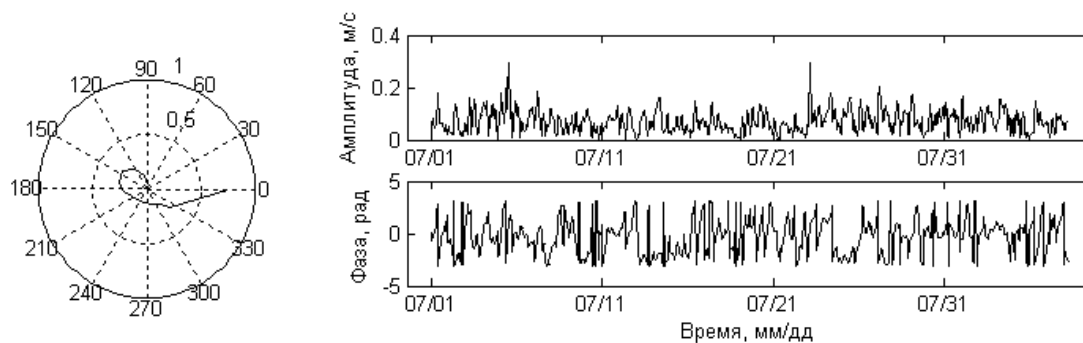


Рис.2. Первая мода и соответствующий модальный коэффициент.

При интерпретации [8] первая мода в анализе поверхностного течения ассоциируется с баротропной составляющей, а первый модальный коэффициент представляет собой траекторию движения частиц жидкости, причем самое масштабное течение осуществляется по часовой стрелке.

В нашем случае самое масштабное течение в соответствии с первым модальным коэффициентом происходит против часовой стрелки (рис.3). На этом рисунке изображен годограф первого модального коэффициента для измерения, которое проводилось прибором ADCP600 в период с 17/06/2014 по 30/07/2014, звездочка соответствует первому измерению по времени.

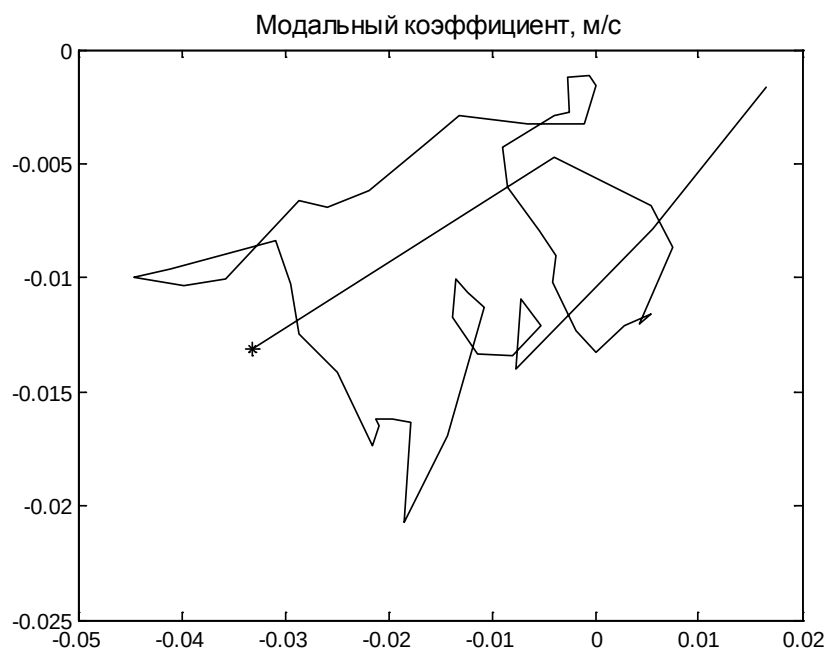


Рис.3. Годограф первого модального коэффициента.

Метод главных компонент был применен к анализу температуры в озере Шира в летний период в десяти точках (рис.4).



Рис.4. Расположение точек длительных измерений температуры в 2015 г.
Рисунок предоставлен с.н.с. ИБ СО РАН Толмеевым А.П.

Проанализируем картину распределения по глубине температуры со временем. На рис.5 изображены вклад в общую энергию всех мод и четыре первые моды.

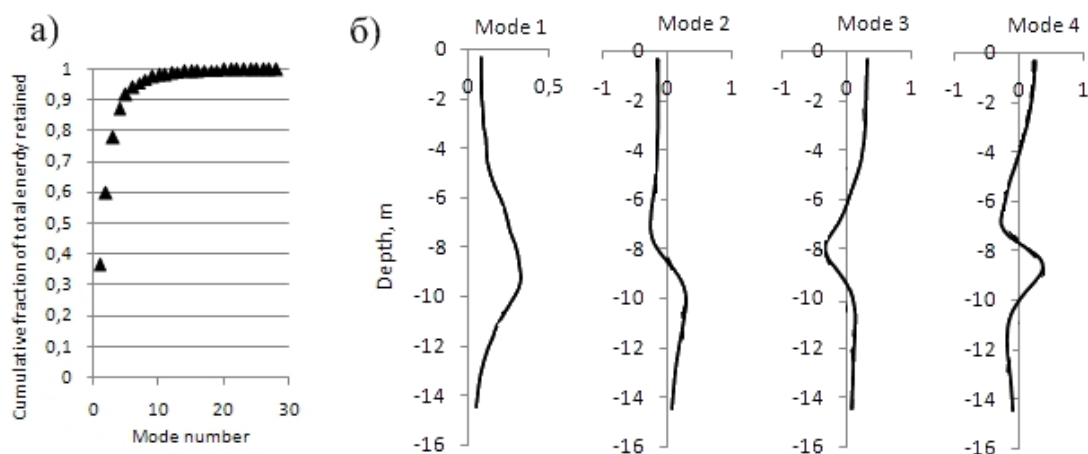


Рис.5. Точка T1, накопленная доля энергии (а) и первые моды (б).

По накопленной энергии видно, что первая мода содержит 30% энергии и четырех мод достаточно для представления в целом поведения температуры (рис.6 а). Наблюдаемая температура соответствует измерению первого июля 2015 г.

Для сравнения укажем, что первая мода для поверхностной температуры в северо-западной части Атлантического океана [11] содержит 80% энергии.

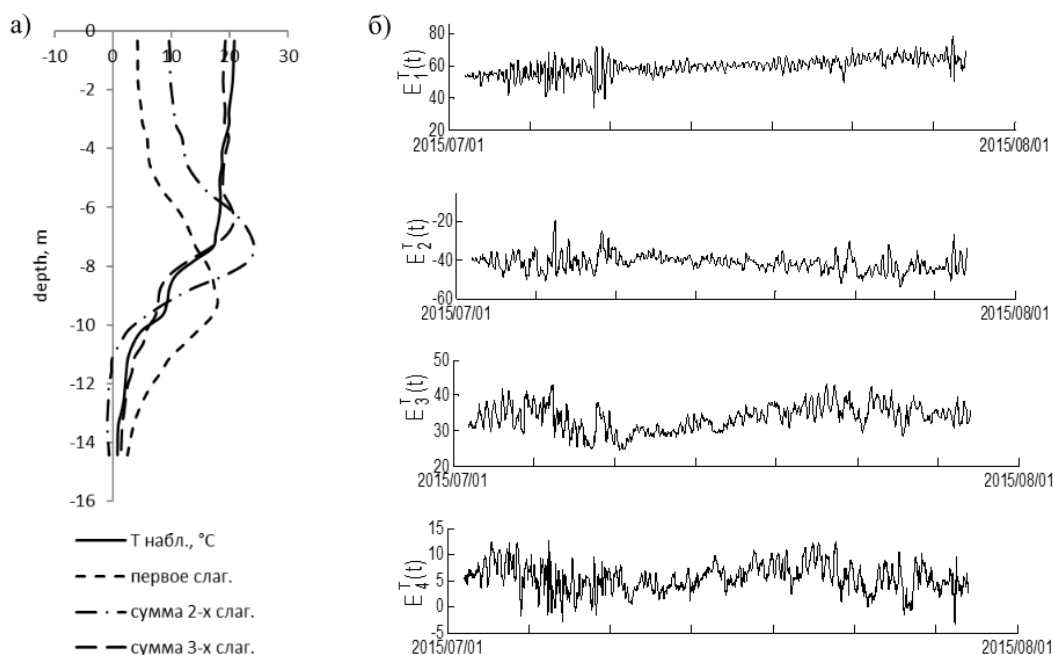


Рис.6. Четыре первые моды (а) и соответствующие им модальные коэффициенты (б).

Далее было проведено сравнение первых мод для семи точек в течение одного периода времени (рис.7).

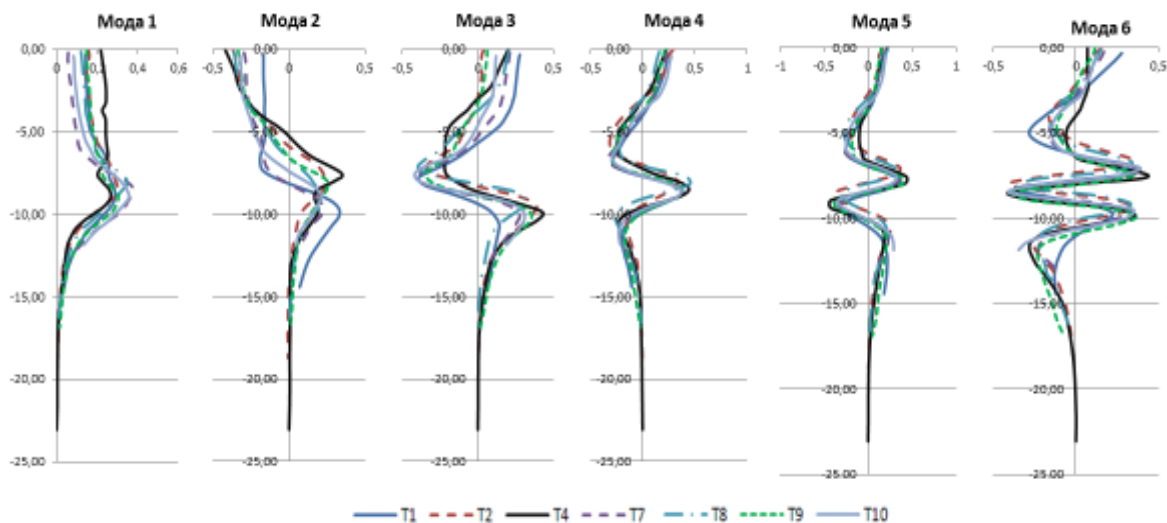
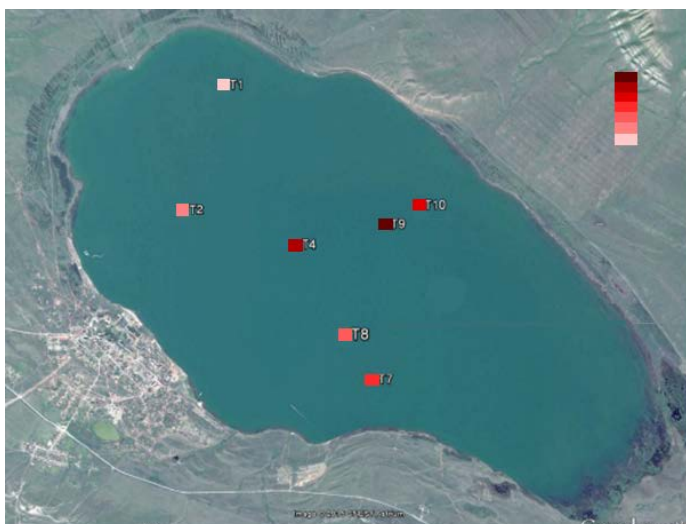


Рис.7. Первые четыре моды для семи точек, в которых измерения проведены на протяжении одного периода времени (27.06.2015 –01.08.2015).

Анализ рис.7 показывает, что первые моды для семи различных точек измерения положительны, что аналогично результату работы [11]. Первые пять мод практически совпадают для всех измерений. Это свидетельствует о том, что вдоль всей акватории процессы, связанные с распределением температуры по глубине, имеют однотипный характер.

При рассмотрении первых от поверхности точек измерения получаем следующую картину распределения температуры на поверхности и первый модальный коэффициент, который соответствует изменению температуры воздуха (рис.8):



T_9	0,4017
T_4	0,3887
T_{10}	0,3868
T_7	0,3760
T_8	0,3714
T_2	0,3676
T_1	0.3515

Рис. 8. Первая мода для поверхностной температуры, аккумулирует 85% энергии.

Сравним поведение первого модального коэффициента для поверхностной температуры с изменением температуры воздуха в период с 27.06.2015 по 01.08.2015 (рис.9).

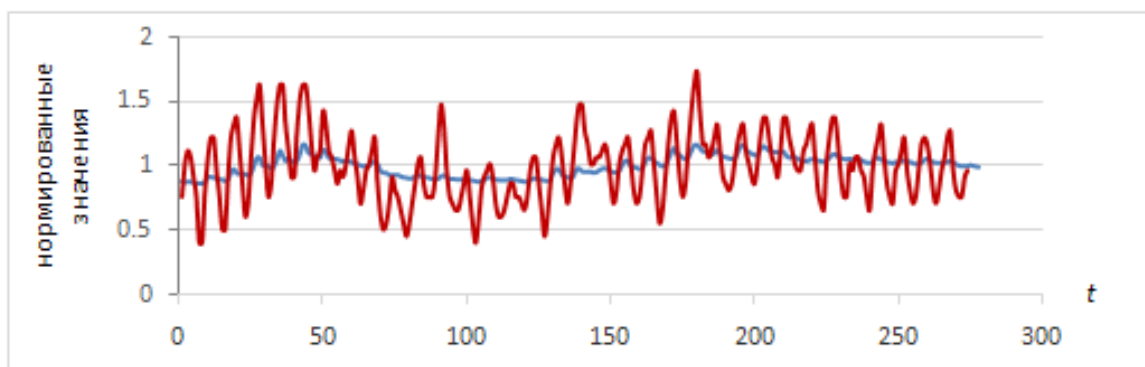


Рис.9. Красный цвет – температура воздуха, отнесенная к среднему значению, синий цвет – первый модальный коэффициент, отнесенный к среднему значению.

При удалении из значений температуры воздуха суточных колебаний методом скользящего среднего по восьми точкам коэффициент корреляции между температурой воздуха и первым модальным коэффициентом равен 0,705, что означает сильную линейную связь между этими величинами.

Список литературы

1. Lorenz E.N. Empirical orthogonal functions and statistical weather prediction / E.N. Lorenz – Sci. Rep. No. 1, Statistical Forecasting Project, M.I.T., Cambridge, MA, 1956. – 48 p.
2. Обухов А.М. О статистически ортогональных разложениях эмпирических функций // Известия АН СССР Серия геофизическая – 1960 – № 3 – С. 432 – 439.
3. Wallace J. M. Empirical orthogonal representation of time series in the frequency domain. Part I: Theoretical considerations / J. M. Wallace, R. E. Dickinson /Journal of Applied Meteorology. – 1972. – V. 11. – №. 6. – p. 887-892.

4. Kaihatu J. M. Empirical orthogonal function analysis of ocean surface currents using complex and real-vector methods/ R. A. Handler, G. O. Marmorino, Shay L. K. //Journal of atmospheric and oceanic technology. – 1998. – V. 15. – №. 4. – p. 927-941.
5. Hannachi A. Empirical orthogonal functions and related techniques in atmospheric science: A review / A. Hannachi, I. T. Jolliffe, D. B. Stephenson //International journal of climatology. – 2007. – V. 27. – №. 9. – p. 1119-1152.
6. Monahan A. H. Empirical orthogonal functions: The medium is the message / A. H. Monahan, J. C. Fyfe, M. H. Ambaum, D. B. Stephenson , G. R. North //Journal of Climate. – 2009. – V. 22. – №. 24. – p. 6501-6514.
7. Alishouse J. C. A discussion of empirical orthogonal functions and their application to vertical temperature profiles / J. C. Alishouse, L. J. Crone, H. E. Fleming, F. L. Van Cleef, D. Q. Wark – Tellus – 1967. – V.19. – № 3. – p. 477-482.
8. Kundu P.K., Allen J.S., Smith R.L., Modal decomposition of the velocity field near the Oregon Coast / P.K. Kundu, J.S. Allen, R.L. Smith // Journal of Physical Oceanography, USA – 1975 – V. 5. - № 4. – p. 683-704.
9. Kundu P. K., Allen J. S. Some three-dimensional characteristics of low-frequency current fluctuations near the Oregon coast / P. K. Kundu, J. S. Allen // Journal of Physical Oceanography. – 1976. – V. 6. – №. 2. – p. 181-199.
10. Kundu P. K. Analysis of current observations on the Georgia shelf / P. K. Kundu, J. O. Blanton, M. M. Janopaul // Journal of Physical Oceanography. – 1981. – V. 11. – №. 8. – p. 1139-1149.
11. Everson R. An empirical eigenfunction analysis of sea surface temperatures in the western North Atlantic / R. Everson, P. Cornillon, L. Sirovich, Webber A. //Journal of Physical Oceanography. – 1997. – V. 27. – №. 3. – p. 468-479.
12. Saurral R. I. Observed modes of sea surface temperature variability in the South Pacific region / R. I. Saurral, F. J. Doblas-Reyes, J. García-Serrano // Climate Dynamics. – 2018. – V. 50. – №. 3-4. – p. 1129-1143.
13. Кузин В.И. Анализ вариаций температуры поверхности тропической и северной частей Тихого океана / В.И. Кузин, А.С. Лобанов // Известия РАН. Физика атмосферы и океана. – 2016. – Т. 52. – № 5 – С. 618-627.
14. Fujii Y. Three-dimensional analysis of temperature and salinity in the equatorial Pacific using a variational method with vertical coupled temperature-salinity empirical orthogonal function modes / Y. Fujii, M. Kamachi. // Journal of Geophysical Research: Oceans. – 2003. – V. 108. – №. C9.
15. Белокопытов В.Н. Статистический анализ термохалинных полей Черного моря по данным буев-профилемеров ARGO / В.Н. Белокопытов, А.В. Багаев. // Экологическая безопасность прибрежной и шельфовой зон и комплексное использование ресурсов шельфа. – Севастополь: ЭКОСИ-Гидрофизика. – 2012. – Вып. 26. – Т. 2. – С. 128 – 142.

ПОСТРОЕНИЕ МОДЕЛИ ОЦЕНКИ НАУЧНОГО ПОТЕНЦИАЛА ИННОВАЦИОННОГО ПРОЕКТА НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

М.А. Габова¹, М.В. Гунер¹, О.И. Пятковский¹, Т.Ю. Солодовникова¹

¹Алтайский государственный технический университет им. И.И. Ползунова,
каф. информационных систем в экономике, gabova_ma@mail.ru

В рамках движения экономики РФ в инновационном русле одним из основных элементов развития науки и техники является их внедрение в производство и введение в практику инновационных технологий. Коммерциализация научных достижений определена одним из приоритетов развития как экономики региона, так и страны в целом.

Реализация инновационных проектов в настоящее время является перспективным направлением, однако для снижения рисков при осуществлении подобной деятельности предпринимателю необходимо в первую очередь провести тщательную оценку предполагаемого к осуществлению проекта. Создание производства наукоемкой продукции, эффективное для одного предприятия, может оказаться неэффективным для другого в силу объективных и субъективных причин, таких, как территориальная расположенность, уровень компетенции персонала по основным направлениям инновационного проекта, состояние основных фондов и т.п. Все эти факторы оказывают влияние на результативность инновационного проекта, но оценить их количественно очень трудно, поэтому еще на стадии отбора необходимо проведение оценки потенциала инновационных проектов. Поскольку на каждом конкретном предприятии существуют свои факторы, влияющие на эффективность инновационных проектов, то универсальной системы оценки проектов нет, но ряд факторов имеет отношение к большинству инновационных предприятий. На основе этих факторов выделяют определенные критерии для оценки инновационных проектов.

В настоящее время подобную оценку осуществляют эксперты, причем для каждого типа проектов создается собственная система критериев. Исходя из вышесказанного видно, что существует необходимость в разработке единой системы критериев, не привязанной к тематической специфике проекта, а также в автоматизации этого процесса. Для решения данного вопроса разрабатывается универсальная адаптивная информационная система оценки инновационного проекта, причем не важно, является ли коммерциализируемый продукт (товар, услуга) научно-техническим, либо представляет собой процесс, новый метод продаж или новый организационный метод в деловой практике, организации рабочих мест или во внешних связях [5].

Для помощи в принятии решения о целесообразности внедрения и поиска путей дальнейшей разработки инновационных проектов создана модель оценки. Разработанная методика состоит из значительного количества факторов, характеризующих потенциал проекта с различных сторон. Эти факторы были объединены в следующие вершины:

1.1 Научный потенциал.

1.2 Предпринимательский потенциал.

1.3 Рыночный потенциал и значимость проекта для экономики и социальной сферы региона.

1.4 Производственный потенциал.

1.5 Кадровый потенциал.

Перечисленные выше факторы являются промежуточными вершинами (узлами) в графе (дереве) решения задачи.

Таким образом, система анализа оценки потенциала научного проекта представляет собой иерархическую модель или ориентированный граф связей, в котором вершинами являются обобщенные оценки, а дугами – зависимости между ними (Рис. 1). При таком представлении решения задачи целесообразно применение гибридных экспертных систем, когда общая задача представляется как совокупность подзадач сформированных в иерархию. Решение производится путем решения всех включенных в нее подзадач.

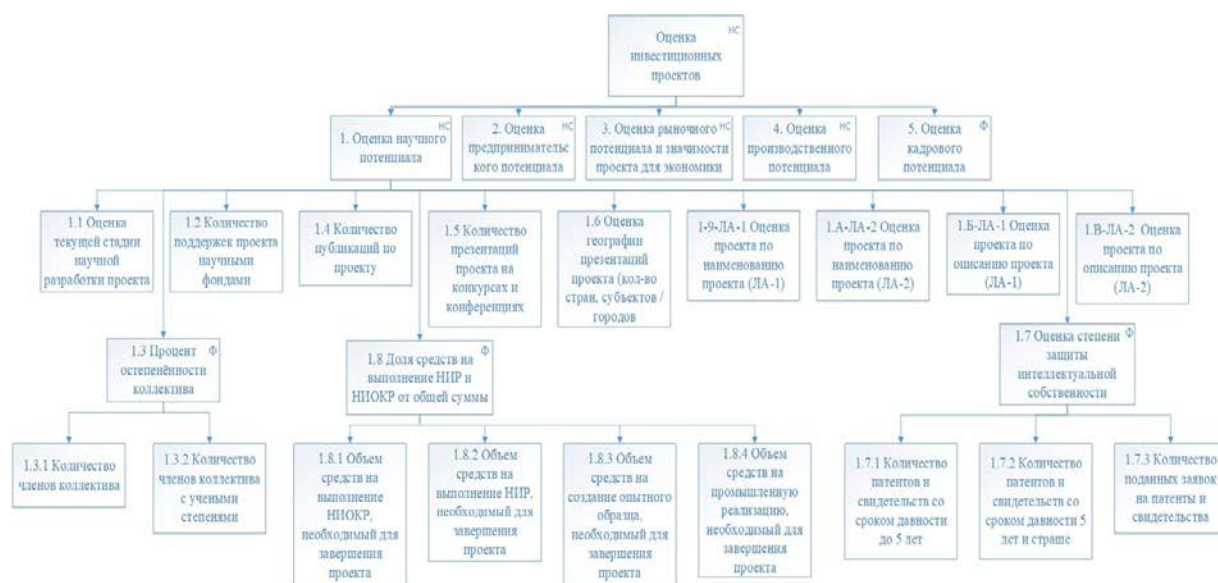


Рисунок 1 – Дерево оценки инновационного потенциала проекта

Созданная экспертная система как реализация гибридной модели представления знаний обеспечивает представление задач в виде иерархического дерева, в котором решение задач вышестоящего уровня базируется на результатах решения задач более низкого уровня. В качестве элементов в систему включены подсистемы или модули, реализующие преобразования F_i предусмотренные гибридной моделью.

Наилучшие результаты, как показывали проводимые эксперименты, при оценке факторов, влияющих на инновационный потенциал, дает нейросетевой анализ. При этом необходимо привлечение группы высоко квалифицированных экспертов, с помощью которых можно создавать полные и непротиворечивые обучающие выборки, при этом применимы методы мозгового штурма, деловых игр, Делфи и др.

В узле «Научный потенциал» характер входных показателей является количественным и непрозрачным. Входные данные со временем также меняются и являются нестабильными, зашумленными и местами не полными, поэтому в данном случае будет целесообразно использовать в качестве метода решения нейронную сеть.

В некоторых узлах рассматриваемого дерева в качестве метода решения определены формализованные зависимости (формулы), так как первичные показатели легко формализуемы и определены, также для показателей, имеющих неколичественную оценку (описательный текст) применялся лингвистический анализ.

Формулы составлены экспертным путем с помощью определения весов каждого показателя в итоговом результате.

Т.е. итоговое значение показателя определяется по следующей формуле:

$$R = \sum_{i=1}^N k_i F_i. \quad (1)$$

где R – результат расчета по формуле, k_i - весовой коэффициент i-го фактора, F_i - значение i-го фактора.

Обучающая выборка сформирована на основе экспертизы проектов, поданных на конкурс инновационных проектов «Новый Алтай» [3].

Таблица 1 - Фрагмент выборки для узла научного потенциал проекта

№ проекта	П-1.1 - Оценка текущей стадии научной разработки проекта - 0/1/2 (3.3)	П-1.2 - Количество подержек проекта научными фондами (3.5)	П-1.3 - Процент острепенности коллектива	П-1.4 - Количество публикаций по проекту (3.7)	П-1.5 - Оценка географии презентации проекта (кол-во стран, субъектов / городов РФ) (3.8)	П-1.6 - Оценка степени защиты интеллектуальной собственности	П-2.6.2, 1.8 - Доля средств на выполнение НИР и НИО КР от общей суммы средств, необходимых для завершения проекта	П-3.1, 2.6.1, 1.9-ЛА-1 - Оценка проекта по наименованию проекта (ЛА-1) (1.1)	П-3.2, 2.6.2, 1.А-ЛА-2 - Оценка проекта по наименованию проекта (ЛА-2) (1.1)	П-3.3, 2.6.3, 1.Б-ЛА-1 - Оценка проекта по описанию проекта (ЛА-1) (2.1)	П-3.4, 2.6.4, 1.В-ЛА-2 - Оценка проекта по описанию проекта (ЛА-2) (2.1)	Оценка научного потенциала
36	0	1	0	4	4	2	0	0	0	0	0	8
38	0	0	0	5	1	0	11	1	0	1	3	7,5
39	0	0	40	0	1	3	0	2	0	3	1	10
40	2	0	0	0	0	0	33	1	1	0	1	3,7
41	0	0	20	4	2	0	1	4	3	3	2	7
42	0	1	0	1	1	2	0	5	1	0	0	10
43	0	0	0	0	2	0	0	1	1	3	0	7,5
44	0	0	40	1	2	2	0	2	1	1	3	9
45	0	0	0	0	7	3	0	2	1	0	1	8,5
46	2	0	0	0	2	0	100	1	1	3	0	8,5
47	2	0	11	3	1	0	24	0	1	0	0	7
49	0	1	0	5	3	0	17	3	1	3	0	7
51	0	0	67	3	2	0	0	3	1	4	0	6,5
52	0	0	0	0	1	0	0	1	0	4	1	7
53	0	1	0	0	4	2	0	1	4	5	3	5
54	0	0	0	0	2	2	0	0	0	0	1	5,5
56	0	1	50	3	4	3	50	3	0	3	2	7
57	0	0	100	0	0	2	0	0	0	1	2	10
59	0	0	0	0	0	0	0	2	2	0	0	8

Для решения задачи оценки инвестиционных проектов в зависимости от различных факторов в аналитической системе "Бизнес Аналитик" была построена соответствующая модель на основе нейронных сетей (рисунок 2). Для автоматизированной загрузки исходных данных и выгрузки рассчитанного по методике

результата были применены web-сервисы "Бизнес Аналитика", доступные в интернете по адресу ba40.ru.

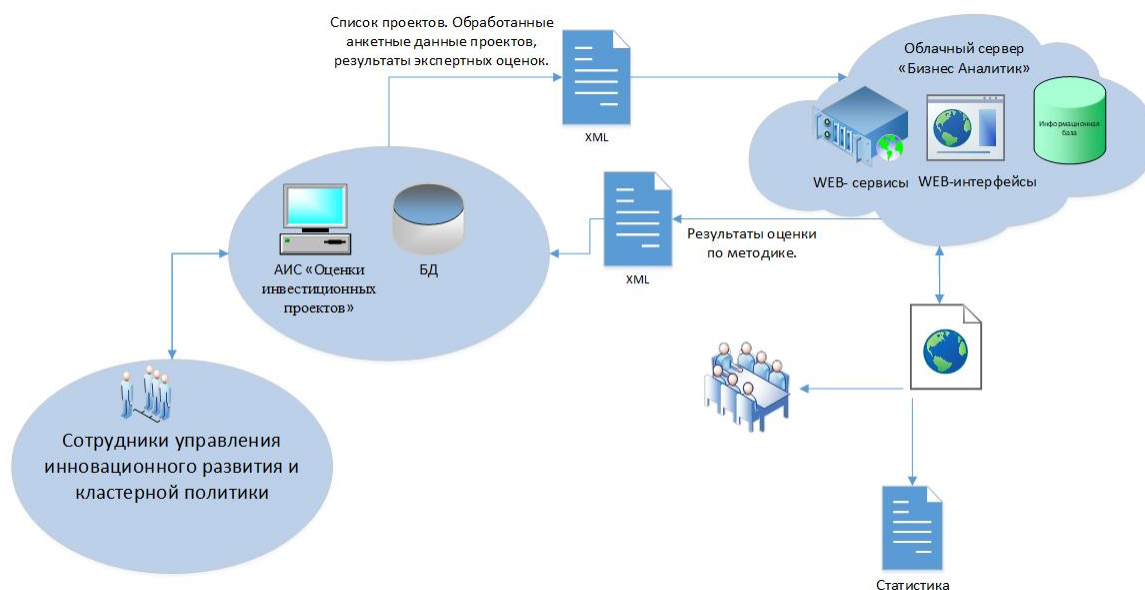


Рисунок 2 – Интеграция с системой «Бизнес-Аналитик»

На основе этих критериев определены методы решения в первом узле дерева – «Научный потенциал», проведено обучение и настройка нейросети. Результаты экспериментов по обучению нейронных сетей различной структуры для узла «Оценка научного потенциала проекта» представлены в таблице 2 и рисунке 3.

Таблица 2 – Структура обучения нейросети в узле «Научный потенциал»

Структура нейронной сети	Ошибка обучения	Ошибка обобщения
5:05:01	0,788	2,004
5:04:01	0,852	2,499
5:03:01	0,846	2,169
5:02:01	0,818	2,208
4:04:01	0,837	2,06
4:03:01	0,768	1,86
4:02:01	0,878	2,156
3:03:01	0,809	2,306
3:02:01	1,033	2,319
2:02:01	0,927	2,6
5:01	0,832	2,395
4:01	0,817	1,572
3:01	0,964	2,417
2:01	1,032	2,045

График зависимостей ошибок обучения и обобщения от структуры нейросети (для узла «Оценка научного потенциала проекта»)

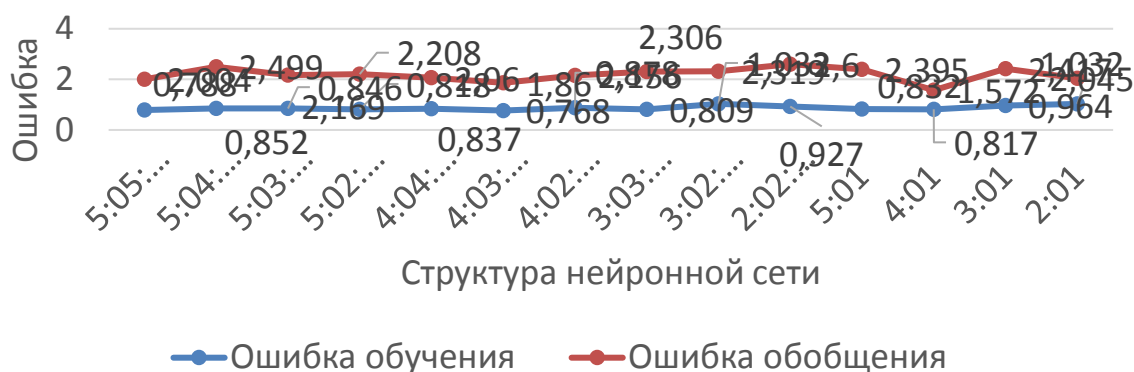


Рисунок 3 – График зависимостей ошибок обучения обобщения от структуры нейросети (для узла «Оценка научного потенциала»)

Программный комплекс обеспечивает гибкость представленной модели за счет средств адаптивной настройки под предметную область. В частности, для нейросетевых решателей используются методы автоматического конструирования и дообучения.

По итогам работы разработанной системы видно, что ее использование является оправданным и целесообразным, так как применяемая система, основанная на экспертизе специалистов дает 78,6% точности, а разработанная система 89,3%, что на 10,7 % выше применяемой в данный момент. Полученный результат позволяет более точно спрогнозировать развитие инновационного проекта в условиях неопределенности, что дает лучший экономический эффект от проведения государственной поддержки наукоемкого бизнеса.

Список литературы:

1. Пятковский О.И. Интеллектуальные компоненты аналитических информационных систем управления организацией. Барнаул: АлтГТУ, 2002. 219 с.
2. Круглов В.В., Дли М.И. Интеллектуальные информационные системы: компьютерная поддержка систем нечеткой логики и нечеткого вывода. М.: Физматлит, 2002.
3. Постановление Администрации Алтайского края от 31.10.2008 № 464 «О ежегодном краевом конкурсе инновационных проектов «Новый Алтай», Указ Губернатора Алтайского края от 09.10.2014г. № 150, Постановление Администрации Алтайского края от 02.09.2016 г. № 308
4. Нейроинформатика / А.Н. Горбань [и др.]. Новосибирск: Наука, Сибирская издат. фирма РАН, 1998. 296 с.
5. Попов Э.В. Экспертные системы: Решение неформализованных задач в диалоге с ЭВМ. М.: Наука, Физматлит, 1987. 288 с
6. Федеральный закон от 21.07.2011 г. № 254-ФЗ «О внесении изменений в федеральный закон «О науке и государственной и научно-технической политике»

НЕКОДИРУЮЩИЕ ОБЛАСТИ ГЕНОМОВ ХЛОРОПЛАСТОВ ОТРАЖАЮТ ТАКСОНОМИЮ ИХ НОСИТЕЛЕЙ

И.К. Горбань², В.С. Густов², М.Ю. Сенашова¹, М.Г. Садовский^{1,2}

¹Институт вычислительного моделирования СО РАН,

²Сибирский федеральный университет, ИФБиТ,

inn.gorban@gmail.com, v.gustows@mail.ru, msen@icm.krasn.ru, msad@icm.krasn.ru

Введение

Одной из важнейших задач современной биологии является выявление и изучение структур в нуклеотидных последовательностях различного уровня и различного таксономического и функционального положения. Здесь интерес вызывает как структурно-функциональный, так и эволюционный аспекты этих отношений. Если эволюционное направление позволяет получить картину развития биологических систем от конкретного вида до глобальных сообществ, то выявление связи между структурой и функцией позволяет раскрыть многие важные механизмы функционирования молекулярно-генетических систем различного уровня и различной природы.

Данная работа направлена на изучение связи структуры и таксономии носителей геномов на примере хлоропластов. Хлоропласты очень удобный объект для исследований, поскольку, во-первых, для них полностью (в хорошем приближении) исключается влияние различий в кодируемых функциях. Размер геномов хлоропластов сравнительно мал ($\approx 10^5$ п.н.), что делает удобным объектом исследований. Поток работ, посвящённых изучению эволюционных процессов и анализу особенностей структуры последовательностей геномов [1, 2] либо отдельных фрагментов и генов [3 – 5] хлоропластов, весьма велик. При этом часто утверждения об эволюции вида делаются на основании анализа изменений генома хлоропластов.

В данной работе мы исследовали связь определённых структур, выявляемых в геномах хлоропластов, с филогенией их носителей. Под структурой мы будем понимать частотный словарь триплетов, определяемый по отдельному формально выделяемому фрагменту генома; точнее, мы будем рассматривать наборы таких словарей, определённые для семейства фрагментов, каждый из которых выделяется в геноме формально. Здесь следует сделать одно важное уточнение: если в работах [7, 8] такие фрагменты выделялись по всему геному, вне зависимости от того, кодирующим был выделяемый фрагмент генома или нет, то в рамках настоящей работы мы сосредоточимся на исследовании связи структуры частот триплетов и таксономии носителей для фрагментов, определяемых исключительно по некодирующим фрагментам генома.

Ранее [7, 8] было показано, что геномы бактерий обладают выраженной семикластерной структурой. Хлоропласты также обладают подобной структурой, однако для геномов хлоропластов выделяется не семь, а восемь кластеров [9 – 11]. Опишем подробнее процедуру построения словарей и выявления структурированности на их множестве. Рассматриваются генетические последовательности длины L , состоящие из символов алфавита $\aleph = \{A, C, G, T\}$. Для каждой из последовательностей составляется частотный словарь толщины 3. Под частотным словарем толщины 3 будем понимать список всех троек $\omega = v_1 v_2 v_3$ идущих подряд нуклеотидов с указанием частот этих троек. Всего может быть 64 триплета. Частота f_ω — это отношение числа копий n_ω данного слова к общему числу всех триплетов N , где N — сумма всех n_ω :

$$f_\omega = \frac{n_\omega}{N}. \quad (1)$$

Триплеты внутри фрагмента определялись следующим образом: они не пересекались, но объединение всех триплетов полностью покрывало всю последовательность во фрагменте. Иными словами, рамка считывания при построении словаря W_3 смещалась на три нуклеотида, а не на один. При этом словарь W_3 задает отображение генома в 64-мерное метрическое пространство. Два генома считаются близкими, если соответствующие им точки в 64-мерном пространстве близки в смысле Евклидовой метрики.

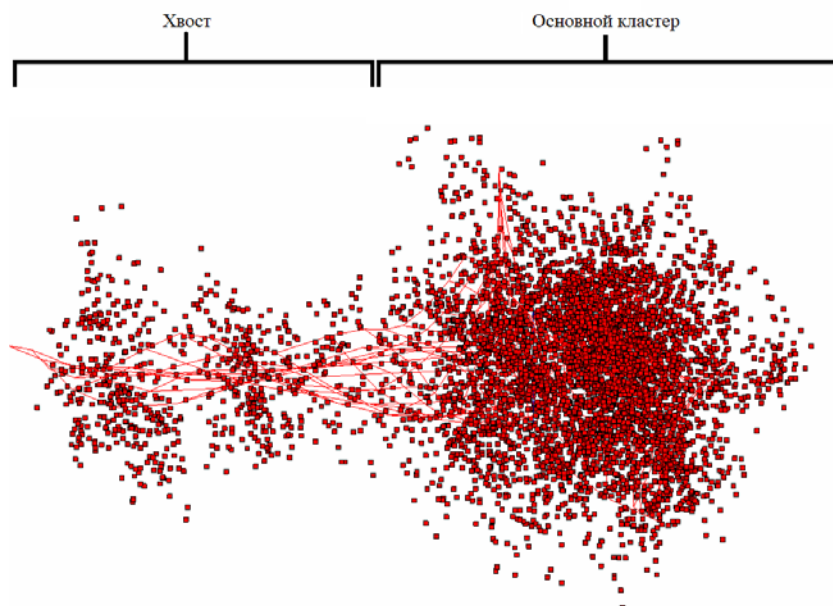


Рис. 1. Два типа фрагментов, выделяемых в структуре некодирующих участков генома, на примере генома *Hordeum vulgare subsp. spontaneum*.

Для построения частотных словарей триплетов последовательность каждого генома сканировалась окном длины Δ с шагом t . Для каждого положения i рамки определялся участок генетической последовательности, совпадающий с рамкой считывания, для которого вычислялся частотный словарь $W_3^{(i)}$, соответствующий i -ой точке в 64-мерном пространстве. С каждой точкой в 64-мерном пространстве связывался номер центрального символа рассматриваемого участка. Номер центрального символа участка совпадает с номером этого символа в последовательности.

Результаты и обсуждение

Для исследования были использованы данные, содержащиеся в открытом доступе в базе EMBL-банка. Был отобран 391 геном хлоропластов различных растений. Для всех генетических последовательностей длина рамки считывания $\Delta = 603$, шаг $t = 11$. По полученному множеству точек строился вид данных в пространстве первых трёх главных компонент, вычисленных для данного 63-мерного пространства в программе *VidaExpert* (<http://bioinfo-out.curie.fr/projects/vidaexpert/>).

Было выделено пять основных структур с двумя типами кластеров, названных **основной кластер** и **хвост**. Для начала опишем кластер типа хвост: это сравнительно небольшой набор фрагментов из некодирующих областей генома, которые обладает специфическими статистическими свойствами, что приводит к его выделению в явно выраженную структурную единицу, по виду действительно напоминающую рыбий хвост. Основной кластер включает в себя все остальные фрагменты и при этом имеет две

различные формы: первая — это шар, вторая — своего рода линза. Эти два типа и кластер типа хвост комбинируются в различных геномах. В Таблице 1 представлены результаты анализа такого рода структур, наблюдаемых в различных геномах.

Таблица. 1. Список исследованных видов отдела Tracheophyta, их таксономия и структура. Условные обозначения структур U: S – шар, S1– шар с одним хвостом, S2 – шар с двумя хвостами, L1 – линза с одним хвостом, L2 – линза с двумя хвостами.

Номер	Вид	Отряд	Класс	U
KP742350	<i>Abies koreana</i>	Pinales	Pinopsida	S1
KF753631	<i>Acer buergerianum subsp. ningpoense</i>	Sapindales	Magnoliopsida	S1
HQ337793	<i>Acidosasa purpurea</i>	Poales	Magnoliopsida	S1
KC844054	<i>Aconitum barbatum var. puberulum</i>	Ranunculales	Magnoliopsida	S1
KF534489	<i>Aegilops cylindrica</i>	Poales	Magnoliopsida	S1
KF534490	<i>Aegilops geniculata</i>	Poales	Magnoliopsida	S1
JQ740834	<i>Aegilops speltoides</i>	Poales	Magnoliopsida	S1
JQ754651	<i>Aegilops tauschii</i>	Poales	Magnoliopsida	S1
JF826503	<i>Ageratina adenophora</i>	Asterales	Magnoliopsida	S1
KF527884	<i>Agrostemma githago</i>	Caryophyllales	Magnoliopsida	S1
KF728079	<i>Allium cepa</i>	Asparagales	Magnoliopsida	S1
KF728080	<i>Allium cepa</i>	Asparagales	Magnoliopsida	L2
KC180806	<i>Allosyncarpia ternata</i>	Myrtales	Magnoliopsida	S1
KF150644	<i>Andrographis paniculata</i>	Lamiales	Magnoliopsida	S1
DQ821119	<i>Angiopteris evecta</i>	Marattiales	Polypodiopsida	S1
KC180805	<i>Angophora costata</i>	Myrtales	Magnoliopsida	L1
KC180804	<i>Angophora floribunda</i>	Myrtales	Magnoliopsida	L1
AP000423	<i>Arabidopsis thaliana</i>	Brassicales	Magnoliopsida	L2
KC456163	<i>Aralia undulata</i>	Apiales	Magnoliopsida	S1
KC465962	<i>Ardisia polysticta</i>	Ericales	Magnoliopsida	S1
JX293720	<i>Artemisia frigida</i>	Asterales	Magnoliopsida	S1
KF887960	<i>Artemisia montana</i>	Asterales	Magnoliopsida	S1
KC817462	<i>Arundinaria appalachiana</i>	Poales	Magnoliopsida	S1
JX513413	<i>Arundinaria fargesii</i>	Poales	Magnoliopsida	S1
JX235347	<i>Arundinaria gigantea</i>	Poales	Magnoliopsida	S1

KF539844	<i>Asclepias nivea</i>	Gentianales	Magnoliopsida	S1
KF386166	<i>Asclepias syriaca</i>	Gentianales	Magnoliopsida	S1
KF986530	<i>Azadirachta indica</i>	Sapindales	Magnoliopsida	S1
HQ337797	<i>Bambusa emeiensis</i>	Poales	Magnoliopsida	S1
KF176554	<i>Berberis bealei</i>	Ranunculales	Magnoliopsida	S2
EU325680	<i>Brachypodium distachyon</i>	Poales	Magnoliopsida	S1
KC456164	<i>Brassaiopsis hainla</i>	Apiales	Magnoliopsida	S1
KF753632	<i>Camellia crapnelliana</i>	Ericales	Magnoliopsida	S1
KF156833	<i>Camellia cuspidata</i>	Ericales	Magnoliopsida	S1
KF156834	<i>Camellia danzaiensis</i>	Ericales	Magnoliopsida	S1
KJ806274	<i>Camellia grandibracteata</i>	Ericales	Magnoliopsida	S1
KF156835	<i>Camellia impressinervis</i>	Ericales	Magnoliopsida	S1
KJ806275	<i>Camellia leptophylla</i>	Ericales	Magnoliopsida	S1
KJ806276	<i>Camellia petelotii</i>	Ericales	Magnoliopsida	S1
KF156837	<i>Camellia pitardii</i>	Ericales	Magnoliopsida	S1
KJ806277	<i>Camellia pubicosta</i>	Ericales	Magnoliopsida	S1
KJ806278	<i>Camellia reticulata</i>	Ericales	Magnoliopsida	S1
KC143082	<i>Camellia sinensis</i>	Ericales	Magnoliopsida	S1
KF156836	<i>Camellia taliensis</i>	Ericales	Magnoliopsida	S1
KF156838	<i>Camellia yunnanensis</i>	Ericales	Magnoliopsida	S1
KP462881	<i>Campynema lineare</i>	Liliales	Magnoliopsida	S1
JX270811	<i>Capsicum annuum</i>	Solanales	Magnoliopsida	S1
KJ619462	<i>Capsicum annuum</i> var. <i>glabriusculum</i>	Solanales	Magnoliopsida	S1
KP274856	<i>Capsicum lycianthoides</i>	Solanales	Magnoliopsida	S1
KP462882	<i>Carludovica palmata</i>	Pandanales	Magnoliopsida	L1
HQ336406	<i>Castanea mollissima</i>	Fagales	Magnoliopsida	S1
KJ001129	<i>Castanopsis echinocarpa</i>	Fagales	Magnoliopsida	S1
KC561139	<i>Catharanthus roseus</i>	Gentianales	Magnoliopsida	S1
KJ690264	<i>Centaurea diffusa</i>	Asterales	Magnoliopsida	S1
KC136217	<i>Cephalotaxus oliveri</i>	Pinales	Pinopsida	S1
JX407108	<i>Ceratozamia hildae</i>	Cycadales	Cycadopsida	S1

JX513415	<i>Chimonocalamus longiusculus</i>	Poales	Magnoliopsida	S1
KP256024	<i>Chloranthus japonicus</i>	Chloranthales	Magnoliopsida	S1
JN867589	<i>Chrysanthemum indicum</i>	Asterales	Magnoliopsida	S1
JN867592	<i>Chrysanthemum indicum</i>	Asterales	Magnoliopsida	S1
JQ362483	<i>Chrysanthemum x morifolium</i>	Asterales	Magnoliopsida	S1
KJ414480	<i>Chrysobalanus icaco</i>	Malpighiales	Magnoliopsida	S1
KC128846	<i>Cistanche deserticola</i>	Lamiales	Magnoliopsida	S1
KJ865401	<i>Citrus aurantiifolia</i>	Sapindales	Magnoliopsida	S1
KM652489	<i>Clematis fusca</i> var. <i>coreana</i>	Ranunculales	Magnoliopsida	S1
JN105689	<i>Colocasia esculenta</i>	Alismatales	Magnoliopsida	L1
JN105690	<i>Colocasia esculenta</i>	Alismatales	Magnoliopsida	L1
JX087681	<i>Corallorhiza striata</i> var. <i>vreelandii</i>	Asparagales	Magnoliopsida	S1
KC180802	<i>Corymbia eximia</i>	Myrtales	Magnoliopsida	S1
KC180800	<i>Corymbia gummifera</i>	Myrtales	Magnoliopsida	S1
KC180801	<i>Corymbia maculata</i>	Myrtales	Magnoliopsida	L1
KC180803	<i>Corymbia tessellaris</i>	Myrtales	Magnoliopsida	L1
KJ414482	<i>Couepia guianensis</i>	Malpighiales	Magnoliopsida	S1
JF412791	<i>Cucumis melo</i> subsp. <i>melo</i>	Cucurbitales	Magnoliopsida	S1
KC427270	<i>Cunninghamia lanceolata</i>	Pinales	Pinopsida	S1
JN867588	<i>Cycas revoluta</i>	Cycadales	Cycadopsida	S1
KC876122	<i>Cymbidium aloifolium</i>	Asparagales	Magnoliopsida	S1
KC876129	<i>Cymbidium mannii</i>	Asparagales	Magnoliopsida	S1
KC876123	<i>Cymbidium sinense</i>	Asparagales	Magnoliopsida	S1
KC876124	<i>Cymbidium tortisepalum</i>	Asparagales	Magnoliopsida	S1
KC876125	<i>Cymbidium tortisepalum</i>	Asparagales	Magnoliopsida	S1
KC876128	<i>Cymbidium tortisepalum</i>	Asparagales	Magnoliopsida	S1
KC876127	<i>Cymbidium tracyanum</i>	Asparagales	Magnoliopsida	S1
KM035764	<i>Cynara cardunculus</i> var. <i>scolymus</i>	Asterales	Magnoliopsida	S1
KJ501998	<i>Cypripedium formosanum</i>	Asparagales	Magnoliopsida	L1
KF925434	<i>Cypripedium macranthos</i>	Asparagales	Magnoliopsida	S1
JN654342	<i>Datura stramonium</i>	Solanales	Magnoliopsida	S1

JN662489	<i>Datura stramonium</i>	Solanales	Magnoliopsida	S1
KC771275	<i>Dendrobium officinale</i>	Asparagales	Magnoliopsida	S1
KP271241	<i>Dendropanax dentiger</i>	Apiales	Magnoliopsida	S1
JN107811	<i>Doroceras hygrometricum</i>	Lamiales	Magnoliopsida	S1
KP308151	<i>Dunalia brachyacantha</i>	Solanales	Magnoliopsida	S1
KP280057	<i>Dunalia obovata</i>	Solanales	Magnoliopsida	S1
KJ000047	<i>Echinochloa crus-galli</i> var. <i>crus-galli</i>	Poales	Magnoliopsida	S1
KJ000048	<i>Echinochloa oryzicola</i>	Poales	Magnoliopsida	S1
JF274081	<i>Elaeis guineensis</i>	Arecales	Magnoliopsida	L1
JN637765	<i>Eleutherococcus senticosus</i>	Apiales	Magnoliopsida	S1
JQ310743	<i>Elodea canadensis</i>	Alismatales	Magnoliopsida	S1
GU191334	<i>Equisetum arvense</i>	Equisetales	Polypodiopsida	S1
KC117177	<i>Equisetum hyemale</i>	Equisetales	Polypodiopsida	S1
KJ523886	<i>Erodium absinthoides</i>	Geraniales	Magnoliopsida	S
HQ713469	<i>Erodium carvifolium</i>	Geraniales	Magnoliopsida	S1
KJ701602	<i>Erodium chrysanthum</i>	Geraniales	Magnoliopsida	S
JF746994	<i>Erycina pusilla</i>	Asparagales	Magnoliopsida	L1
KC180789	<i>Eucalyptus aromaphloia</i>	Myrtales	Magnoliopsida	S1
KC180773	<i>Eucalyptus baxteri</i>	Myrtales	Magnoliopsida	S1
KC180791	<i>Eucalyptus camaldulensis</i>	Myrtales	Magnoliopsida	S1
KC180786	<i>Eucalyptus cladocalyx</i>	Myrtales	Magnoliopsida	S1
KC180779	<i>Eucalyptus cloeziana</i>	Myrtales	Magnoliopsida	S1
KC180782	<i>Eucalyptus curtisii</i>	Myrtales	Magnoliopsida	S1
KC180792	<i>Eucalyptus deglupta</i>	Myrtales	Magnoliopsida	S1
KC180771	<i>Eucalyptus delegatensis</i>	Myrtales	Magnoliopsida	S1
KC180795	<i>Eucalyptus diversicolor</i>	Myrtales	Magnoliopsida	S1
KC180774	<i>Eucalyptus diversifolia</i>	Myrtales	Magnoliopsida	S1
KC180776	<i>Eucalyptus elata</i>	Myrtales	Magnoliopsida	S1
KC180799	<i>Eucalyptus erythrocorys</i>	Myrtales	Magnoliopsida	S1
KC180787	<i>Eucalyptus globulus</i>	Myrtales	Magnoliopsida	S1
KC180798	<i>Eucalyptus guilfoylei</i>	Myrtales	Magnoliopsida	S1

KC180781	<i>Eucalyptus marginata</i>	Myrtales	Magnoliopsida	S1
KC180783	<i>Eucalyptus melliodora</i>	Myrtales	Magnoliopsida	S1
KC180797	<i>Eucalyptus microcorys</i>	Myrtales	Magnoliopsida	S1
KC180788	<i>Eucalyptus nitens</i>	Myrtales	Magnoliopsida	S1
KC180769	<i>Eucalyptus obliqua</i>	Myrtales	Magnoliopsida	S1
KC180780	<i>Eucalyptus patens</i>	Myrtales	Magnoliopsida	S1
KC180785	<i>Eucalyptus polybractea</i>	Myrtales	Magnoliopsida	S1
KC180770	<i>Eucalyptus radiata</i>	Myrtales	Magnoliopsida	S1
KC180777	<i>Eucalyptus regnans</i>	Myrtales	Magnoliopsida	S1
KC180790	<i>Eucalyptus saligna</i>	Myrtales	Magnoliopsida	S1
KC180796	<i>Eucalyptus salmonophloia</i>	Myrtales	Magnoliopsida	S1
KC180775	<i>Eucalyptus sieberi</i>	Myrtales	Magnoliopsida	S1
KC180793	<i>Eucalyptus spathulata</i>	Myrtales	Magnoliopsida	S1
KC180794	<i>Eucalyptus torquata</i>	Myrtales	Magnoliopsida	S1
KC180778	<i>Eucalyptus umbra</i>	Myrtales	Magnoliopsida	S1
KC180772	<i>Eucalyptus verrucata</i>	Myrtales	Magnoliopsida	S1
EU254477	<i>Fagopyrum esculentum</i> subsp. <i>ancestrale</i>	Caryophyllales	Magnoliopsida	L2
JX513416	<i>Fargesia nitida</i>	Poales	Magnoliopsida	L2
JX513417	<i>Fargesia spathacea</i>	Poales	Magnoliopsida	S1
JX513418	<i>Fargesia yunnanensis</i>	Poales	Magnoliopsida	S1
JN884816	<i>Fragaria chiloensis</i>	Rosales	Magnoliopsida	L2
JQ396171	<i>Fragaria vesca</i> subsp. <i>bracteata</i>	Rosales	Magnoliopsida	S1
JN884817	<i>Fragaria virginiana</i>	Rosales	Magnoliopsida	S1
KF769143	<i>Fritillaria cirrhosa</i>	Liliales	Magnoliopsida	L1
KF712486	<i>Fritillaria hupehensis</i>	Liliales	Magnoliopsida	L2
KC543997	<i>Fritillaria taipaiensis</i>	Liliales	Magnoliopsida	S1
JX513419	<i>Gaoligongshania megalothyrsa</i>	Poales	Magnoliopsida	L2
HG530134	<i>Genlisea margaretae</i>	Lamiales	Magnoliopsida	S2
AB684440	<i>Ginkgo biloba</i>	Ginkgoales	Ginkgoopsida	S1
JN654343	<i>Ginkgo biloba</i>	Ginkgoales	Ginkgoopsida	S1

JN867578	<i>Ginkgo biloba</i>	Ginkgoales	Ginkgoopsida	S1
JN867583	<i>Ginkgo biloba</i>	Ginkgoales	Ginkgoopsida	S1
JN867585	<i>Ginkgo biloba</i>	Ginkgoales	Ginkgoopsida	S1
KC893635	<i>Glycine canescens</i>	Fabales	Magnoliopsida	S1
KC893632	<i>Glycine cyrtoloba</i>	Fabales	Magnoliopsida	S1
KC893636	<i>Glycine dolichocarpa</i>	Fabales	Magnoliopsida	S1
KC893637	<i>Glycine falcata</i>	Fabales	Magnoliopsida	S1
DQ317523	<i>Glycine max</i>	Fabales	Magnoliopsida	L2
KF611800	<i>Glycine soja</i>	Fabales	Magnoliopsida	S1
KC893634	<i>Glycine stenophita</i>	Fabales	Magnoliopsida	S1
KC893638	<i>Glycine syndetika</i>	Fabales	Magnoliopsida	S1
KC893633	<i>Glycine tomentella</i>	Fabales	Magnoliopsida	S1
KF201590	<i>Glycyrrhiza glabra</i>	Fabales	Magnoliopsida	L1
KC427271	<i>Gnetum montanum</i>	Ephedrales	Gnetopsida	S2
KJ501999	<i>Goodyera fumata</i>	Asparagales	Magnoliopsida	L2
JF317351	<i>Gossypium anomalum</i>	Malvales	Magnoliopsida	L1
JN019795	<i>Gossypium areysianum</i>	Malvales	Magnoliopsida	L1
HQ901198	<i>Gossypium barbadense</i>	Malvales	Magnoliopsida	L1
HQ901199	<i>Gossypium barbadense</i>	Malvales	Magnoliopsida	L1
HQ901200	<i>Gossypium barbadense</i>	Malvales	Magnoliopsida	L1
JF317352	<i>Gossypium bickii</i>	Malvales	Magnoliopsida	L1
JN019794	<i>Gossypium capitis-viridis</i>	Malvales	Magnoliopsida	L1
HQ901195	<i>Gossypium gossypoides</i>	Malvales	Magnoliopsida	L1
JF317353	<i>Gossypium herbaceum</i>	Malvales	Magnoliopsida	L1
HQ901196	<i>Gossypium hirsutum</i>	Malvales	Magnoliopsida	L1
HQ901197	<i>Gossypium hirsutum</i>	Malvales	Magnoliopsida	L1
JN019792	<i>Gossypium incanum</i>	Malvales	Magnoliopsida	L1
JF317354	<i>Gossypium longicalyx</i>	Malvales	Magnoliopsida	L1
JN019793	<i>Gossypium somalense</i>	Malvales	Magnoliopsida	L1
JF317355	<i>Gossypium stocksii</i>	Malvales	Magnoliopsida	S1
JF317356	<i>Gossypium sturtianum</i>	Malvales	Magnoliopsida	L1

JQ742090	<i>Gossypium turneri</i>	Malvales	Magnoliopsida	L1
EU549769	<i>Guizotia abyssinica</i>	Asterales	Magnoliopsida	S2
KJ524104	<i>Habenaria pantlingiana</i>	Asparagales	Magnoliopsida	L2
DQ383815	<i>Helianthus annuus</i>	Asterales	Magnoliopsida	S2
KP688069	<i>Hibiscus syriacus</i>	Malvales	Magnoliopsida	L2
KJ414485	<i>Hirtella physophora</i>	Malpighiales	Magnoliopsida	S1
KJ414479	<i>Hirtella racemosa</i>	Malpighiales	Magnoliopsida	S1
KC912688	<i>Hordeum vulgare subsp. spontaneum</i>	Poales	Magnoliopsida	S1
KC912687	<i>Hordeum vulgare subsp. vulgare</i>	Poales	Magnoliopsida	S2
AY660566	<i>Huperzia lucidula</i>	Lycopodiales	Lycopodiopsida	S2
KF248009	<i>Hyoscyamus niger</i>	Solanales	Magnoliopsida	L2
KF240616	<i>Hypseocharis bilobata</i>	Geraniales	Magnoliopsida	L2
EF380354	<i>Illicium oligandrum</i>	Austrobaileyales	Magnoliopsida	S2
JX513421	<i>Indocalamus wilsonii</i>	Poales	Magnoliopsida	S1
JX513422	<i>Indosasa sinica</i>	Poales	Magnoliopsida	L2
KP294386	<i>Iochroma nitidum</i>	Solanales	Magnoliopsida	S2
KP212149	<i>Ipomoea batatas</i>	Solanales	Magnoliopsida	S2
HQ234669	<i>Jacobaea vulgaris</i>	Asterales	Magnoliopsida	S2
DQ673255	<i>Jasminum nudiflorum</i>	Lamiales	Magnoliopsida	S2
FJ695500	<i>Jatropha curcas</i>	Malpighiales	Magnoliopsida	L2
KF866297	<i>Juniperus bermudiana</i>	Pinales	Pinopsida	S1
KF866297	<i>Juniperus bermudiana</i>	Pinales	Pinopsida	S1
KF866298	<i>Juniperus monosperma</i>	Pinales	Pinopsida	S1
KF866299	<i>Juniperus scopulorum</i>	Pinales	Pinopsida	S
KF866300	<i>Juniperus virginiana</i>	Pinales	Pinopsida	S1
KC456167	<i>Kalopanax septemlobus</i>	Apiales	Magnoliopsida	S2
AP010820	<i>Keteleeria davidiana</i>	Pinales	Pinopsida	S1
AB501189	<i>Larix decidua</i>	Pinales	Pinopsida	S1
HM029371	<i>Lathyrus sativus</i>	Fabales	Magnoliopsida	S1
DQ400350	<i>Lemna minor</i>	Alismatales	Magnoliopsida	L2
KJ414483	<i>Licania alba</i>	Malpighiales	Magnoliopsida	L2

KJ414481	<i>Licania heteromorpha</i>	Malpighiales	Magnoliopsida	S1
KJ414484	<i>Licania sprucei</i>	Malpighiales	Magnoliopsida	S1
HG530133	<i>Lindenbergia philippensis</i>	Lamiales	Magnoliopsida	L2
KC588388	<i>Liquidambar formosana</i>	Saxifragales	Magnoliopsida	L2
DQ899947	<i>Liriodendron tulipifera</i>	Magnoliales	Magnoliopsida	S2
AM777385	<i>Lolium perenne</i>	Poales	Magnoliopsida	L2
KJ170923	<i>Lonicera japonica</i>	Dipsacales	Magnoliopsida	S2
AP002983	<i>Lotus japonicus</i>	Malpighiales	Magnoliopsida	L2
KC695666	<i>Lupinus luteus</i>	Fabales	Magnoliopsida	L2
KM233640	<i>Luzuriaga radicans</i>	Liliales	Magnoliopsida	L1
KC536645	<i>Lygodium japonicum</i>	Schizaeales	Polypodiopsida	S2
KF862711	<i>Macadamia integrifolia</i>	Proteales	Magnoliopsida	S1
JN227740	<i>Magnolia denudata</i>	Magnoliales	Magnoliopsida	S1
JN867577	<i>Magnolia denudata</i>	Magnoliales	Magnoliopsida	S1
JN867584	<i>Magnolia grandiflora</i>	Magnoliales	Magnoliopsida	S1
JN867587	<i>Magnolia grandiflora</i>	Magnoliales	Magnoliopsida	S1
HM775382	<i>Magnolia kwangsiensis</i>	Magnoliales	Magnoliopsida	S2
JN867579	<i>Magnolia officinalis</i>	Magnoliales	Magnoliopsida	S1
JN867580	<i>Magnolia officinalis subsp. biloba</i>	Magnoliales	Magnoliopsida	S1
JN867581	<i>Magnolia officinalis subsp. Biloba</i>	Magnoliales	Magnoliopsida	S1
JN867582	<i>Magnolia officinalis subsp. biloba</i>	Magnoliales	Magnoliopsida	S1
KF753638	<i>Magnolia yunnanensis</i>	Magnoliales	Magnoliopsida	S1
EU117376	<i>Manihot esculenta</i>	Malpighiales	Magnoliopsida	L2
JF343520	<i>Mankyua chejuensis</i>	Ophioglossales	Polypodiopsida	S2
KP205432	<i>Masdevallia coccinea</i>	Asparagales	Magnoliopsida	L2
KJ566305	<i>Masdevallia picturata</i>	Asparagales	Magnoliopsida	S1
KF241982	<i>Medicago truncatula f. tricycla</i>	Fabales	Magnoliopsida	S1
FJ597983	<i>Megaleranthis saniculifolia</i>	Ranunculales	Magnoliopsida	L2
KC456165	<i>Metapanax delavayi</i>	Apiales	Magnoliopsida	S2
JN673818	<i>Millettia pinnata</i>	Fabales	Magnoliopsida	L2
DQ226511	<i>Morus indica</i>	Rosales	Magnoliopsida	L2

KM491711	<i>Morus mongolica</i>	Rosales	Magnoliopsida	L1
AB830885	<i>Nageia nagi</i>	Pinales	Pinopsida	S1
JX978472	<i>Najas flexilis</i>	Alismatales	Magnoliopsida	L2
AP009376	<i>Nasturtium officinale</i>	Brassicales	Magnoliopsida	L2
JQ336992	<i>Nelumbo lutea</i>	Proteales	Magnoliopsida	L1
JQ336993	<i>Nelumbo nucifera</i>	Proteales	Magnoliopsida	L1
KF009944	<i>Nelumbo nucifera</i>	Proteales	Magnoliopsida	L1
KM655836	<i>Nelumbo nucifera</i>	Proteales	Magnoliopsida	L1
KF356392	<i>Neyraudia reynaudiana</i>	Poales	Magnoliopsida	L2
AB237912	<i>Nicotiana sylvestris</i>	Solanales	Magnoliopsida	L2
JN563929	<i>Nicotiana undulata</i>	Solanales	Magnoliopsida	S1
DQ354691	<i>Nuphar advena</i>	Nymphaeales	Magnoliopsida	L2
AJ627251	<i>Nymphaea alba</i>	Nymphaeales	Magnoliopsida	L2
KF753633	<i>Nymphaea mexicana</i>	Nymphaeales	Magnoliopsida	S1
EU262887	<i>Oenothera argillicola</i>	Myrtales	Magnoliopsida	S2
JX513423	<i>Oligostachyum shiuyingianum</i>	Poales	Magnoliopsida	L2
HG803179	<i>Orobanche gracilis</i>	Lamiales	Magnoliopsida	S2
AY522329	<i>Oryza sativa Indica Group</i>	Poales	Magnoliopsida	S2
JN861109	<i>Oryza sativa Indica Group</i>	Poales	Magnoliopsida	L2
JN861110	<i>Oryza sativa Indica Group</i>	Poales	Magnoliopsida	L2
AY522330	<i>Oryza sativa Japonica Group</i>	Poales	Magnoliopsida	S2
GU592207	<i>Oryza sativa Japonica Group</i>	Poales	Magnoliopsida	S2
X15901	<i>Oryza sativa Japonica Group</i>	Poales	Magnoliopsida	L2
JQ806762	<i>Pachycladon cheesemanii</i>	Brassicales	Magnoliopsida	L2
JX205495	<i>Pachycladon ensyii</i>	Brassicales	Magnoliopsida	S1
KJ206533	<i>Paeonia obovata</i>	Saxifragales	Magnoliopsida	S2
KF753636	<i>Paeonia sp. Sd0052</i>	Saxifragales	Magnoliopsida	S1
AY582139	<i>Panax ginseng</i>	Apiales	Magnoliopsida	S2
KF431956	<i>Panax ginseng</i>	Apiales	Magnoliopsida	S1
HQ731441	<i>Panicum virgatum</i>	Poales	Magnoliopsida	L2
HQ822121	<i>Panicum virgatum</i>	Poales	Magnoliopsida	S1

KJ566307	<i>Paphiopedilum armeniacum</i>	Asparagales	Magnoliopsida	L2
KJ524105	<i>Paphiopedilum niveum</i>	Asparagales	Magnoliopsida	L1
KJ414486	<i>Parinari campestris</i>	Malpighiales	Magnoliopsida	L2
GU120098	<i>Parthenium argentatum</i>	Asterales	Magnoliopsida	S2
KF240617	<i>Pelargonium alternans</i>	Geraniales	Magnoliopsida	S1
DQ897681	<i>Pelargonium x hortorum</i>	Geraniales	Magnoliopsida	S2
JQ041763	<i>Pentactina rupicola</i>	Rosales	Magnoliopsida	L2
JX436155	<i>Penthorum chinense</i>	Saxifragales	Magnoliopsida	L2
JF719062	<i>Phalaenopsis equestris</i>	Asparagales	Magnoliopsida	L2
KC311467	<i>Pharus lappulaceus</i>	Poales	Magnoliopsida	L2
KF954541	<i>Phedimus takesimensis</i>	Saxifragales	Magnoliopsida	S1
KF730315	<i>Phragmites australis</i>	Poales	Magnoliopsida	L2
HQ337796	<i>Phyllostachys edulis</i>	Poales	Magnoliopsida	L2
JN415113	<i>Phyllostachys propinqua</i>	Poales	Magnoliopsida	L2
HF937082	<i>Picea abies</i>	Pinales	Pinopsida	S1
HG803178	<i>Pinguicula ehlersiae</i>	Lamiales	Magnoliopsida	S2
EU998740	<i>Pinus contorta</i>	Pinales	Pinopsida	S1
KC427272	<i>Pinus massoniana</i>	Pinales	Pinopsida	S1
KC427273	<i>Pinus taeda</i>	Pinales	Pinopsida	S1
DQ887677	<i>Piper cenocladum</i>	Piperales	Magnoliopsida	S2
HM029370	<i>Pisum sativum</i>	Fabales	Magnoliopsida	S1
JX513424	<i>Pleiolblastus maculatus</i>	Poales	Magnoliopsida	L2
KJ010812	<i>Podocarpus lambertii</i>	Pinales	Pinopsida	S1
KC306742	<i>Podocarpus totara</i>	Pinales	Pinopsida	S
KJ664927	<i>Populus balsamifera</i>	Malpighiales	Magnoliopsida	S1
KJ624919	<i>Populus euphratica</i>	Malpighiales	Magnoliopsida	S1
KJ664926	<i>Populus fremontii</i>	Malpighiales	Magnoliopsida	S1
KF922320	<i>Praxelis clematidea</i>	Asterales	Magnoliopsida	S1
KM981744	<i>Premna microphylla</i>	Lamiales	Magnoliopsida	S2
KF753634	<i>Primula poissonii</i>	Ericales	Magnoliopsida	S2
KF990036	<i>Prunus kansuensis</i>	Rosales	Magnoliopsida	L2

KP760071	<i>Prunus maximowiczii</i>	Rosales	Magnoliopsida	S1
KF765450	<i>Prunus mume</i>	Rosales	Magnoliopsida	L1
KP760072	<i>Prunus padus</i>	Rosales	Magnoliopsida	L1
HQ336405	<i>Prunus persica</i>	Rosales	Magnoliopsida	L1
KP760073	<i>Prunus serrulata</i> var. <i>spontanea</i>	Rosales	Magnoliopsida	L1
KP760074	<i>Prunus subhirtella</i> var. <i>subhirtella</i>	Rosales	Magnoliopsida	L1
KP760075	<i>Prunus subhirtella</i> var. <i>subhirtella</i>	Rosales	Magnoliopsida	L1
KP760070	<i>Prunus yedoensis</i>	Rosales	Magnoliopsida	L1
AB601120	<i>Pseudotsuga sinensis</i>	Pinales	Pinopsida	S1
AP004638	<i>Psilotum nudum</i>	Psilotales	Polypodiopsida	S2
KC534841	<i>Puelia olyriiformis</i>	Poales	Magnoliopsida	L2
KP301144	<i>Quercus aliena</i>	Fagales	Magnoliopsida	L2
KP340971	<i>Quercus aquifolioides</i>	Fagales	Magnoliopsida	S1
KM841421	<i>Quercus spinosa</i>	Fagales	Magnoliopsida	S1
DQ359689	<i>Ranunculus macranthus</i>	Ranunculales	Magnoliopsida	L2
KJ716483	<i>Raphanus sativus</i>	Brassicales	Magnoliopsida	L2
KJ617081	<i>Retrophyllum piresii</i>	Pinales	Pinopsida	S1
KJ123753	<i>Rhazya stricta</i>	Gentianales	Magnoliopsida	L2
JF937588	<i>Ricinus communis</i>	Malpighiales	Magnoliopsida	L1
KF753637	<i>Rosa odorata</i> var. <i>gigantea</i>	Rosales	Magnoliopsida	L2
AP006714	<i>Saccharum hybrid</i>	Poales	Magnoliopsida	L2
KJ742926	<i>Salix interior</i>	Malpighiales	Magnoliopsida	L2
HF586694	<i>Salvia miltiorrhiza</i>	Lamiales	Magnoliopsida	L2
JX312195	<i>Salvia miltiorrhiza</i>	Lamiales	Magnoliopsida	S1
JX513414	<i>Sarocalamus faberi</i>	Poales	Magnoliopsida	L2
KC456166	<i>Schefflera delavayi</i>	Apiales	Magnoliopsida	S2
HG738866	<i>Schwalbea americana</i>	Lamiales	Magnoliopsida	S2
JX427551	<i>Sedum sarmentosum</i>	Saxifragales	Magnoliopsida	L2
JN637766	<i>Sesamum indicum</i>	Lamiales	Magnoliopsida	S2
KC569603	<i>Sesamum indicum</i>	Lamiales	Magnoliopsida	S1
KJ001642	<i>Setaria italica</i>	Poales	Magnoliopsida	L2

KF527886	<i>Silene chalcedonica</i>	Caryophyllales	Magnoliopsida	L2
JF715054	<i>Silene conica</i>	Caryophyllales	Magnoliopsida	S1
KF527885	<i>Silene conoidea</i>	Caryophyllales	Magnoliopsida	S1
JF715055	<i>Silene latifolia</i>	Caryophyllales	Magnoliopsida	S1
JF715056	<i>Silene noctiflora</i>	Caryophyllales	Magnoliopsida	S1
KF527887	<i>Silene paradoxa</i>	Caryophyllales	Magnoliopsida	S1
JF715057	<i>Silene vulgaris</i>	Caryophyllales	Magnoliopsida	S1
JF772170	<i>Solanum tuberosum</i>	Solanales	Magnoliopsida	S1
JF772171	<i>Solanum tuberosum</i>	Solanales	Magnoliopsida	S1
EF115542	<i>Sorghum bicolor</i>	Poales	Magnoliopsida	L2
JN160603	<i>Spirodela polyrhiza</i>	Alismatales	Magnoliopsida	L2
JX416858	<i>Stangeria eriopus</i>	Cycadales	Cycadopsida	S2
KC180807	<i>Stockwellia quadrifida</i>	Myrtales	Magnoliopsida	L2
KC427274	<i>Taiwania flousiana</i>	Pinales	Pinopsida	S1
JN867586	<i>Taxus mairei</i>	Pinales	Pinopsida	S
JN867590	<i>Taxus mairei</i>	Pinales	Pinopsida	S
JN867591	<i>Taxus mairei</i>	Pinales	Pinopsida	S
KJ123824	<i>Taxus mairei</i>	Pinales	Pinopsida	S1
KC608752	<i>Tetracentron sinense</i>	Trochodendrales	Magnoliopsida	L2
JX513425	<i>Thamnocalamus spathiflorus</i>	Poales	Magnoliopsida	L2
EU090187	<i>Trachelium caeruleum</i>	Asterales	Magnoliopsida	S2
KJ476730	<i>Trifolium meduseum</i>	Fabales	Magnoliopsida	S1
EU849487	<i>Trifolium subterraneum</i>	Fabales	Magnoliopsida	S1
KF990556	<i>Trigonobalanus doichangensis</i>	Fagales	Magnoliopsida	S2
AB042240	<i>Triticum aestivum</i>	Poales	Magnoliopsida	L2
KC912694	<i>Triticum aestivum</i>	Poales	Magnoliopsida	S1
KJ592713	<i>Triticum aestivum</i>	Poales	Magnoliopsida	L2
KC912690	<i>Triticum monococcum</i>	Poales	Magnoliopsida	S1
KM352501	<i>Triticum turgidum subsp. durum</i>	Poales	Magnoliopsida	S1
KC912693	<i>Triticum urartu</i>	Poales	Magnoliopsida	S1
KC608753	<i>Trochodendron aralioides</i>	Trochodendrales	Magnoliopsida	L1

KC997777	<i>Utricularia gibba</i>	Lamiales	Magnoliopsida	S2
JQ248601	<i>Vaccinium macrocarpon</i>	Ericales	Magnoliopsida	S2
KJ566306	<i>Vanilla planifolia</i>	Asparagales	Magnoliopsida	S1
KF437397	<i>Veratrum patulum</i>	Liliales	Magnoliopsida	L2
GQ893027	<i>Vigna radiata</i>	Fabales	Magnoliopsida	S2
JQ755301	<i>Vigna unguiculata</i>	Fabales	Magnoliopsida	S1
KF976463	<i>Vitis rotundifolia</i>	Vitales	Magnoliopsida	L2
DQ424856	<i>Vitis vinifera</i>	Vitales	Magnoliopsida	L2
AB856289	<i>Vitis vinifera subsp. Caucasica</i>	Vitales	Magnoliopsida	L2
AB856290	<i>Vitis vinifera subsp. Caucasica</i>	Vitales	Magnoliopsida	L2
AB856291	<i>Vitis vinifera subsp. Caucasica</i>	Vitales	Magnoliopsida	L2
KF240615	<i>Viviania marifolia</i>	Geraniales	Magnoliopsida	S2
EU342371	<i>Welwitschia mirabilis</i>	Ephedrales	Gnetopsida	S2
JN160605	<i>Wolffia australiana</i>	Alismatales	Magnoliopsida	L2
JN160604	<i>Wolffiella lingulata</i>	Alismatales	Magnoliopsida	L1
JX513426	<i>Yushania levigata</i>	Poales	Magnoliopsida	L2
JX416857	<i>Zamia furfuracea</i>	Cycadales	Cycadopsida	S2
X86563	<i>Zea mays</i>	Poales	Magnoliopsida	L2
KF241980	<i>Zea mays subsp. mays</i>	Poales	Magnoliopsida	L2
KF241981	<i>Zea mays subsp. mays</i>	Poales	Magnoliopsida	L2

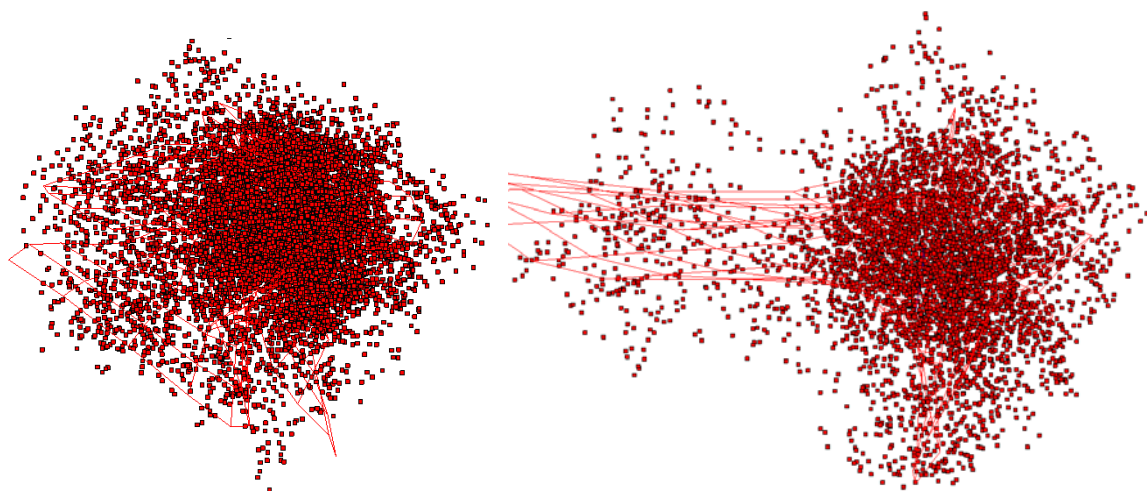


Рис. 2. Структура типа «Шар» на примере генома *Erodium chrysanthum* (слева) и типа «Шар с одним хвостом» на примере генома *Pseudotsuga sinensis* (справа).

Кроме представленных в таблице видов, 6 растений относились к другим отделам:

- Три вида растений (*Aneura mirabilis*, AC EU043314, *Ptilidium pulcherrimum*, AC HM222519 и *Apopellia endiviifolia*, AC JX827163) относились к отделу *Marchantiophyta* и все имели структуру «Шар с двумя хвостами».
- Два вида растений (*Tetraphis pellucida*, AC KJ817846 и *Nyholmiella obtusifolia*, AC KP765733) относились к отделу *Bryophyta* и оба имели структуру «Шар с одним хвостом»
- Один вид растений (*Anthoceros angustus*, AC AB086179) относился к отделу *Anthocerotophyta* и имел структуре «Шар с двумя хвостами».

Рис. 2 – 4 показывают упомянутые выше структуры.

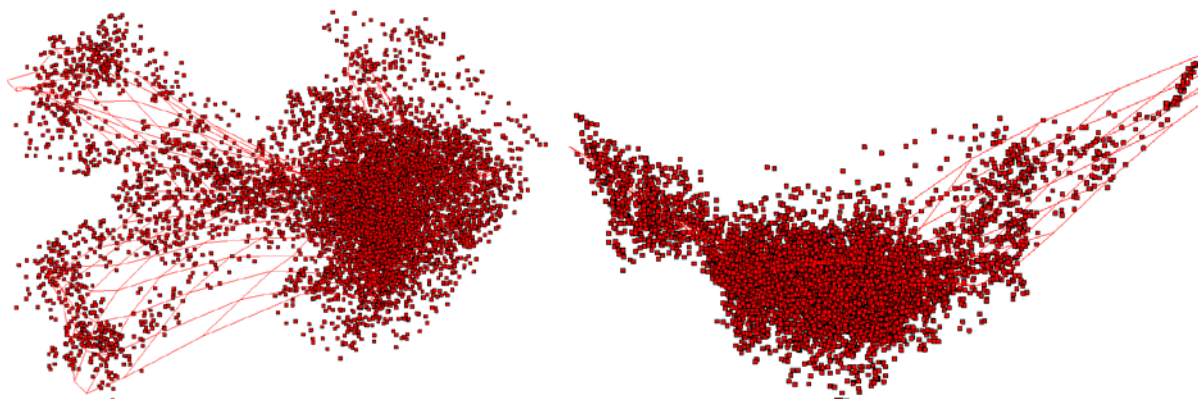


Рис. 3. Структура типа «Шар с двумя хвостами» на примере генома *Liriodendron tulipifera* (слева) и типа «Линза с одним хвостом» на примере генома *Ricinus communis* (справа).

Следует отметить, что кластер **хвост** всегда находится левее основного кластера. Данное обстоятельство можно отнести на счёт произвола в выборе направления первой главной компоненты, однако выбор её направления всегда проводился системно, исключительно на основе статистических свойств частотных словарей, соответствующих формально выделяемым фрагментам. Из этого, в частности, следует, что структура некодирующих частей генома хлоропластов является по-своему консервативной: по крайней мере, взаимное расположение групп статистически заметно различающихся фрагментов, скорее всего, совпадает в геномах разных видов.

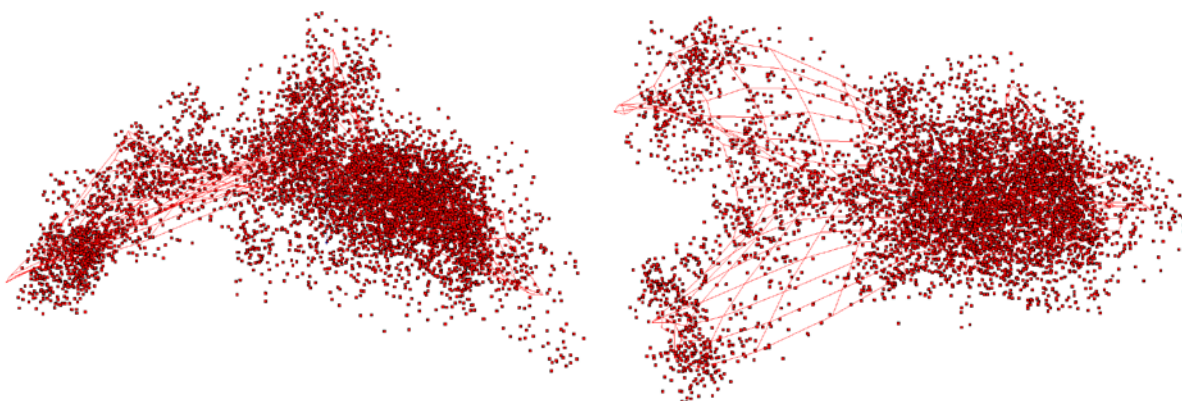


Рис. 4. Структура типа «Линза с двумя хвостами» на примере генома *Lupinus luteus*, в двух проекциях.

Структура типа «Шар» характеризуется шарообразным видом *основного кластера* и отсутствием *хвоста*. Структура типа «Шар с одним хвостом»

характеризуется шарообразным видом *основного кластера* и наличием одного *хвоста*. Структура типа «Шар с двумя хвостами» характеризуется шарообразным видом *основного кластера* и наличием двух, чётко обособленных *хвостов*. Структура типа «Линза с одним хвостом» характеризуется сочетанием *основного кластера* и *хвоста*, в совокупности напоминающим полумесяц, обращенный дистальной, относительно центра, стороной вверх или вниз. Эта структура проявляется лишь в проекции типа «вид сверху». В основной проекции структура выглядит как шар с одним *хвостом*. Структура «Линза с двумя хвостами» характеризуется сочетанием *основного кластера* и *хвоста*, в совокупности напоминающим полумесяц, обращенный дистальной, относительно центра, стороной вверх или вниз. Аналогично, этот вид проявляется лишь в проекции «сверху», в основной же проекции структура выглядит как шар с двумя *хвостами*.

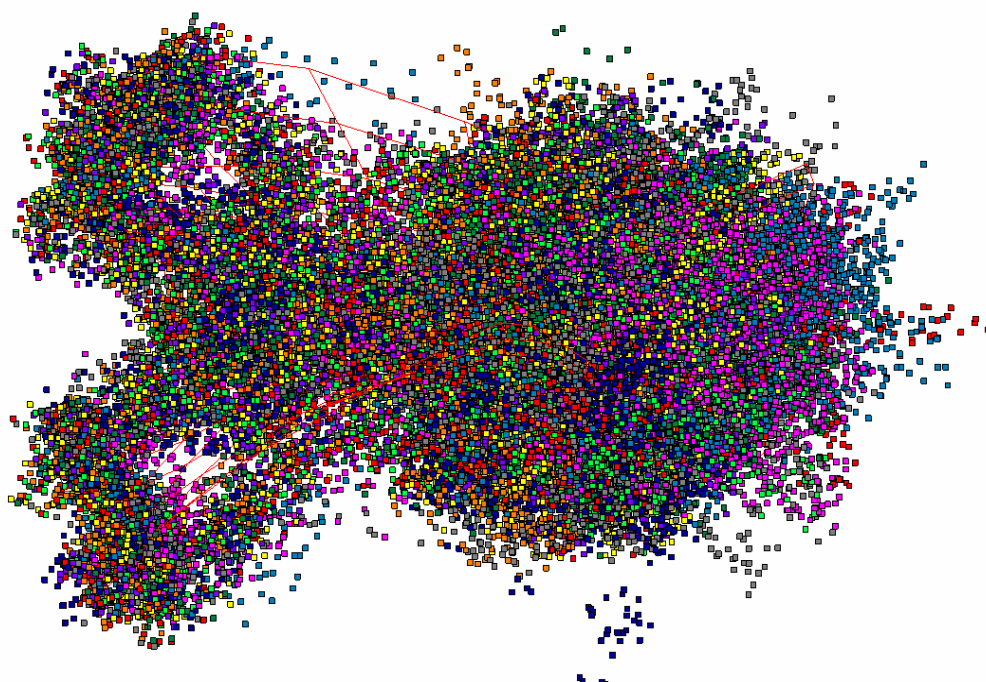


Рис. 5. Объединённая структура геномов *Genlisea margaretae*, *Schwalbea americana*, *Pinguicula ehlersiae*, *Orobanchе gracilis*, *Ptilidium pulcherrimum*, *Magnolia kwangsiensis*, *Jacobaea vulgaris*, *Mankyua chejuensis*, *Sesamum indicum* и *Vaccinium macrocarpon*.

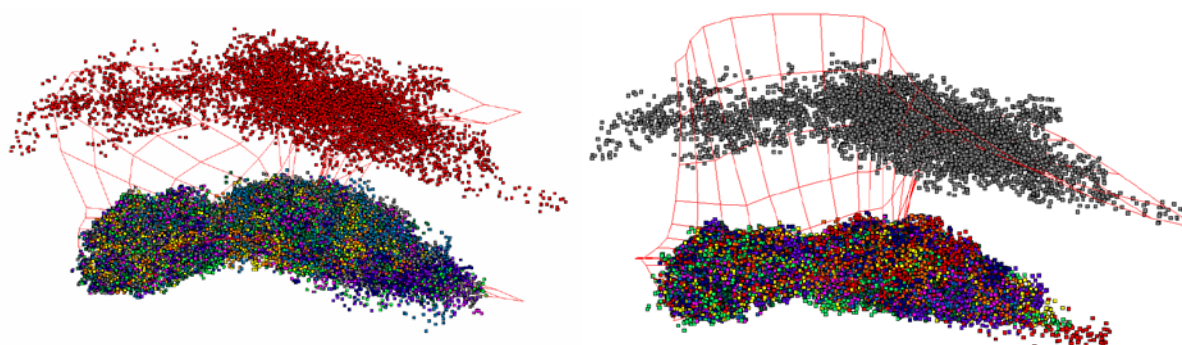


Рис. 6. Некодирующий геном хлоропластов *Allium cepa* (окрашен в красный цвет, KF728079) выделяется на фоне не кодирующих геномов хлоропластов других растений. Рисунки отличаются разными наборами геномов, формирующих общий кластер.

Перейдём теперь к обсуждению тех структур, которые можно выделить в ансамблях формально выделяемых фрагментов, принадлежащих разным видам, в случае одновременной кластеризации таких геномов. Для того, чтобы показать, что целостность структуры проявляется не только в каждом отдельно взятом геноме, но и в их множестве, была построена общая структура для 10 геномов. Множество точек, принадлежащее каждому из геномов, окрашивалась в отдельный цвет (см. Рис.5). В большинстве случаев объединение некодирующих участков 10 произвольно взятых геномов хлоропластов, разделённых на формально выделяемые фрагменты, как описано выше, порождали структуру практически идентичную между собой; типичный пример такой структурированности показан на Рис.5.

Однако среди всех геномов выделяется группа из геномов 9 видов, которые ведут себя принципиально иначе, в пространстве частот триплетов, определяемым по формально выделяемым фрагментам в некодирующих областях геномов; это поведение показано на Рис.6. Поясним сказанное. Для этого обратимся к Рис.5. На этом рисунке хорошо видно, что фрагменты некодирующих участков геномов различных организмов имеют ярко выраженную тенденцию группироваться во вполне определённых местах той общей структуры типа «шар с двумя кольцами», что отчётливо видно по расположению различно окрашенных точек.

При этом очевидно, что никакими методами кластеризации выделить индивидуальные подгруппы фрагментов, принадлежащие только одному виду, не удастся. Не вызывает сомнения тот факт, что кластеризация структуры типа той, что представлена на Рис.5 методом динамических ядер приведёт к построению какой-то классификации. Эта классификация может оказаться устойчивой или неустойчивой, классы, выделяемые ею, могут оказаться различимыми или неразличимыми, но в любом случае ожидать того, что в один класс попадут точки, соответствующие фрагментам одного вида, не приходится.

Противоположная картина наблюдается для ряда организмов; пример представлен на Рис.6. Было выявлено 9 видов растений, геномы хлоропластов которых формируют ярко выраженный отдельный кластер от остальных геномов. Они принадлежали таким видам как: *Allium cepa*, *Angiopteris evecta*, *Anthoceros angustus*, *Brachypodium distachyon*, *Cycas revoluta*, *Equisetum arvense*, *Ginkgo biloba*, *Larix decidua*, *Pseudotsuga sinensis*. Для всех этих видов показано, что их формально выделяемые фрагменты формируют отдельный кластер, который наверняка будет выделять любым методом кластеризации — от метода динамических ядер до топологических методов. Обращает на себя внимание также тот факт, что в этом списке видов присутствуют виды самых разных таксонов высокого уровня. Причины столь необычного поведения этих видов требуют отдельного изучения.

Заключение

Для всех рассмотренных в данной статье геномов хлоропластов установлено, что структура распределения формально выделяемых коротких фрагментов из некодирующих областей генома в пространстве первых трёх главных компонент, определяемых для частотных словарей триплетов, очень похожа: имеется центральное ядро и один или два «хвоста». Всего выделяется пять различных структур, отличающихся формой распределения, а также наличием или отсутствием кластера «хвост». Все эти структуры выделяются при индивидуальном анализе геномов, каждого вида отдельно. Анализ структуры, получаемой при совместном распределении формально выделяемых фрагментов, представляющих несколько разных видов, показывает, что большая часть этих видов формирует структуры, которые не могут быть разделены на кластеры со значимым преобладанием фрагментов от генома одного вида;

точнее, группирование фрагментов одного и того же вида легко идентифицируется в такой структуре, однако оно не может быть выделено методами кластеризации. При этом девять видов полностью выделяются из указанной схемы: при построении совместного с ними распределения фрагменты этих видов полностью обособляются от остальных и формируют кластер, который может быть устойчиво и различимо выделен любыми методами кластеризации.

Список литературы

1. Leliaert F., Smith D.R., Moreau H., Herron M.D., Verbruggen H., Delwiche Ch.F., De Clerck O., 2012. Phylogeny and Molecular Evolution of the Green Algae // *Critical Reviews in Plant Sciences*. V. 31. P. 1-46.
2. Carbonell-Caballero J., Alonso R., Ibanez V., Terol J., Talon M., Dopazo J., 2015. A phylogenetic analysis of 34 chloroplast genomes elucidates the relationships between wild and domestic species within the genus citrus // *Molecular Biology and Evolution*. V. 32. № 8. P. 2015--2035.
3. Milanowski R., Zakrys B., Kwiatowski J., 2001. Phylogenetic analysis of chloroplast smallsubunit rRNA genes of the genus *Euglena* Ehrenberg // *International Journal of Systematic and Evolutionary Microbiology*. V. 51. P. 773-781.
4. Dong W. , Liu J., Yu J., Wang L., Zhou Sh., 2012. Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding // *PLoS ONE*. V. 7. № 4. P. 1-9.
5. Marazzi B., Endress P.K., De Queiroz L.P., Conti E., 2006. Phylogenetic relationships within senna (*Leguminosae*, *Cassiinae*) based on three chloroplast dna regions: patterns in the evolution of floral symmetry and extrafloral nectaries // *American Journal of Botany*. V. 93. № 2. P. 288-303.
6. Bugaenko N. N., Gorban A.N., Sadovsky M. G. (1998) Maximum entropy method in analysis of genetic text and measurement of its information content. *Open Systems & Information Dyn.* 5(2): 265-278.
7. Gorban A. N., Zinovyev A. Yu., Popova T. G. (2003) Seven clusters in genomic triplet distributions. *In Silico Biology* 3: 39-45.
8. Gorban A. N., Zinovyev A. Yu., Popova T. G. (2005) Universal seven-cluster structure of genome fragment distribution: basic symmetry in triplet frequencies. *Bioinformatics of Genome Regulation and Structure II*. (Eds. N. Kolchanov and R. Hofstaedt) Springer Science+Business Media, Inc. P. 153-163.
9. Садовский М.Г., Сенашова М.Ю., Малышев А.В. Восьмикластерная структура геномов хлоропластов наземных растений // *ЖОБ*. 2018, т.79(2), С.124 – 134.
10. Садовский М.Г., Сенашова М.Ю., Малышев А.В. Chloroplast Genomes Exhibit Eight-Cluster Structuredness and Mirror Symmetry // *LNBI*, vol.10813, 2018, pp.186 – 196.
11. Садовский М.Г., Сенашова М.Ю., Путинцева Ю.А. Eight Clusters, Synchrony of Evolution and Unique Symmetry in Chloroplast Genomes: The Offering from Triplets. В: *Chloroplasts and Cytoplasm: Structure and Functions* (ser. Cell Biology Research Progress), Eds. Carl Dejesus and Lourdes Trask, 2018, Nova Science Publishers, Inc., N.-Y., pp.25 – 95.

РАЗРАБОТКА МОДЕЛЕЙ И СИСТЕМЫ ОЦЕНКИ И ПРОГНОЗА СПРОСА НА ПАРФЮМЕРНУЮ ПРОДУКЦИЮ НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

М.В. Гунер¹, Р.Е. Ткаченко¹, О.И. Пятковский¹

¹Алтайский государственный технический университет им. И.И. Ползунова,
каф. информационных систем в экономике, horyzont1@mail.ru

По данным "TNS Россия", за последние десять лет женщины в России стали использовать косметику на 13,3% больше. Кроме того, среднестатистическая женщина в России потребляет значительно больше декоративной косметики, чем среднестатистическая женщина в Европе. Особенно высокий спрос на парфюмерную и косметическую продукцию наблюдается в декабре и жаркие летние месяцы. А в преддверии Нового года и праздников также меняется соотношение объемов продаж косметики и парфюмерии различных классов: бизнес-, эконом-, люкс- и среднего класса.

Объектом исследования является ООО «Парфюм Вояж» — крупнейшая в Барнауле сеть парфюмерно-косметических магазинов. Компания уже 14 лет занимается развитием индустрии красоты.

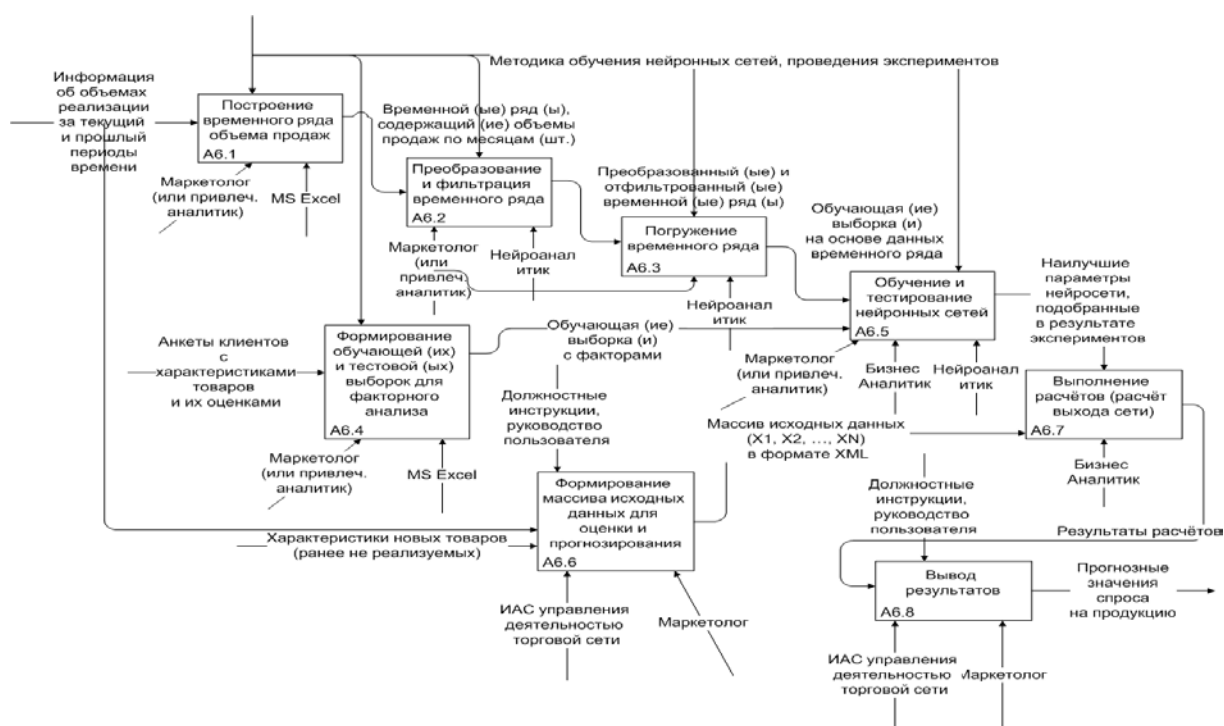


Рис. 1. Описание процесса оценки и прогноза спроса на продукцию «как должно быть»

Цель работы – построить модель прогноза спроса на парфюмерную продукцию на основе исторических данных об объемах продаж, построить модель оценки спроса на парфюмерную продукцию на основе факторного анализа, спроектировать и реализовать информационную систему для автоматизации процедур оценки и прогноза.

Для решения задач оценки и прогноза спроса были выбраны нейронные сети ввиду их отличных аппроксимирующих способностей, возможности извлечения знаний из имеющихся массивов данных [1, 2]. В качестве программных средств для проведения вычислительных экспериментов были использованы системы «Нейро-аналитик» и

«Бизнес Аналитик», разработанные на кафедре информационных систем в экономике Алтайского государственного технического университета им. И.И. Ползунова. Процесс оценки и прогноза спроса на продукцию «как должно быть» описан на рис. 1.

Модель прогноза спроса на парфюмерную продукцию на основе исторических данных об объемах продаж.

В качестве исходных данных были взяты помесечные объемы продаж парфюмерной продукции (в штуках) за период с 01.01.2013 по 01.02.2018 (см. таблицу 1).

Таблица 1

Фрагмент временных рядов с данными об объемах продаж
парфюмерной продукции различных классов

Дата	Объем продаж парфюмерной продукции (шт.)			
	Люкс-класс	Бизнес-класс	Средний класс	Эконом-класс
01.01.2013	170	605	605	190
01.02.2013	393	929	1161	536
01.03.2013	353	882	1148	328
01.04.2013	201	515	656	248
01.05.2013	175	534	745	220
01.06.2013	248	675	757	286
01.07.2013	264	609	609	729
01.08.2013	318	771	929	283
01.09.2013	388	787	1038	257
01.10.2013	266	679	893	309
01.11.2013	348	876	925	323
01.12.2013	760	2141	2464	697
01.01.2014	183	484	632	191
01.02.2014	845	1769	2021	633
01.03.2014	482	1194	1455	463
...				
01.03.2017	202	524	862	514
01.04.2017	59	254	402	280

01.05.2017	63	266	406	276
01.06.2017	71	306	480	353
01.07.2017	122	390	578	384
01.08.2017	123	491	701	444
01.09.2017	74	363	521	396
01.10.2017	119	442	652	371
01.11.2017	121	405	607	338
01.12.2017	411	843	1670	767
01.01.2018	94	241	416	215
01.02.2018	117	448	821	443

Исходные данные, используемые для обучения нейронных сетей – объемы продаж за период с 01.01.2013 по 01.09.2017; для тестирования обученных сетей (собственно прогноза) и оценки адекватности модели – объемы продаж за период с 01.10.2017 по 01.02.2018 (см. рис. 2).

Метод фильтрации временных рядов – экспоненциальное среднее скользящее. Метод обучения нейронных сетей – kParTan. Структура сети – 4:1.

Покажем результаты проведенных вычислительных экспериментов на примере парфюмерной продукции бизнес-класса (см. рис. 3). Наименьшую ошибку прогнозирования 9,9% мы получили при глубине погружения 18. Средняя ошибка прогнозирования на 3 периода вперед зачастую больше, чем на 5 периодов вперед, т.к. 3-ий прогнозируемый период – это декабрь, который всегда характеризуется очень высокими, порой даже аномально высокими, продажами по сравнению со среднемесячными.

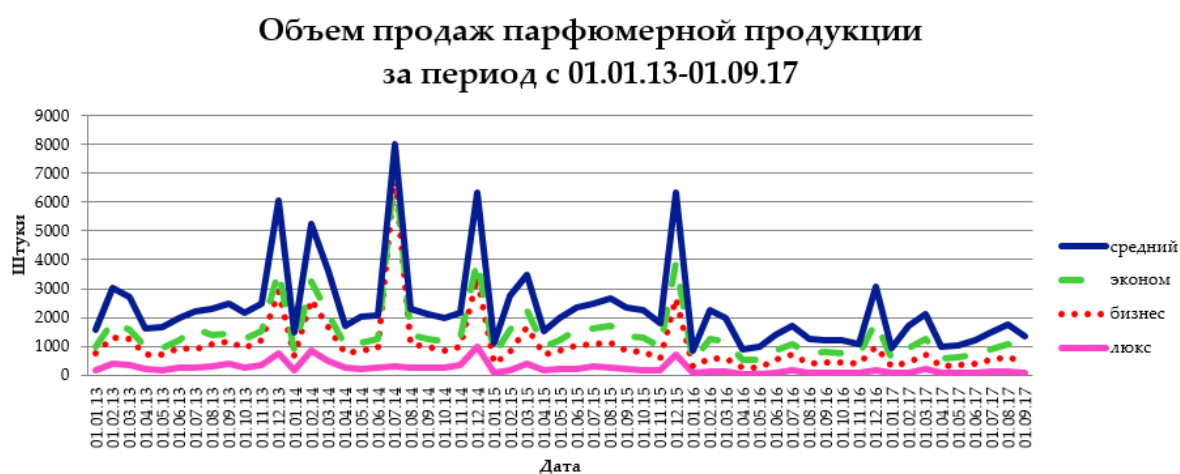


Рис. 2. Динамика объемов продаж парфюмерной продукции
за период с 01.01.2013-01.09.2017

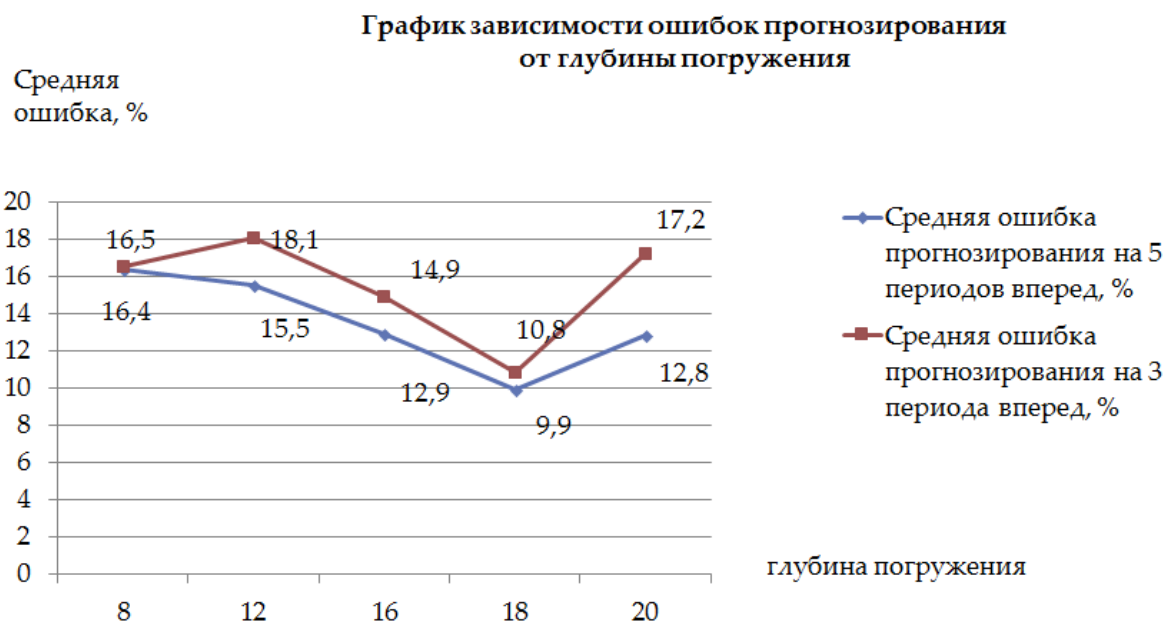


Рис. 3. Результаты экспериментов по обучению и тестированию нейронных сетей
(на примере парфюмерной продукции бизнес-класса)

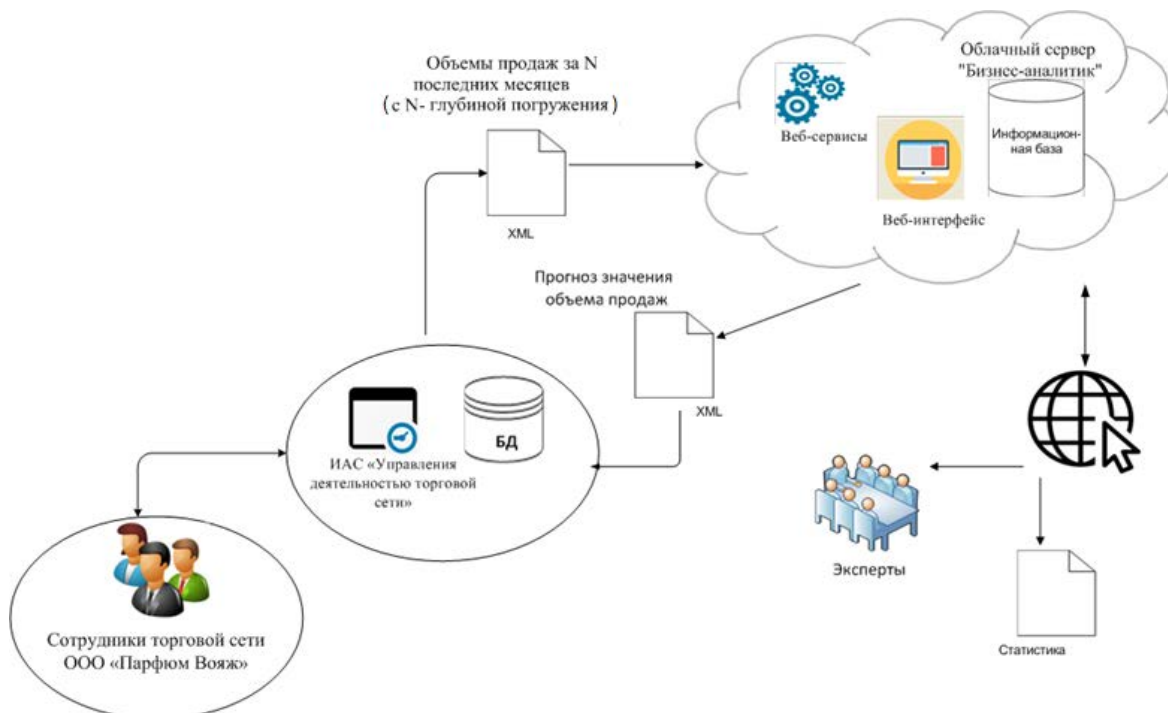


Рис. 4. Схема интеграции информационно-аналитической системы управления деятельностью торговой сети с WEB-сервисами системы «Бизнес Аналитик»

С целью автоматизации процедуры прогнозирования спроса на парфюмерную продукцию и соответственно объемов продаж были использованы WEB-сервисы «Бизнес Аналитика», доступные в интернете по адресу ba40.ru. В качестве формата

обмена данных между системой «Бизнес Аналитик» и информационно-аналитической системой управления деятельностью торговой сети, реализованной на платформе «1С: Предприятие 8», использовался формат XML (см. рис. 4-6).

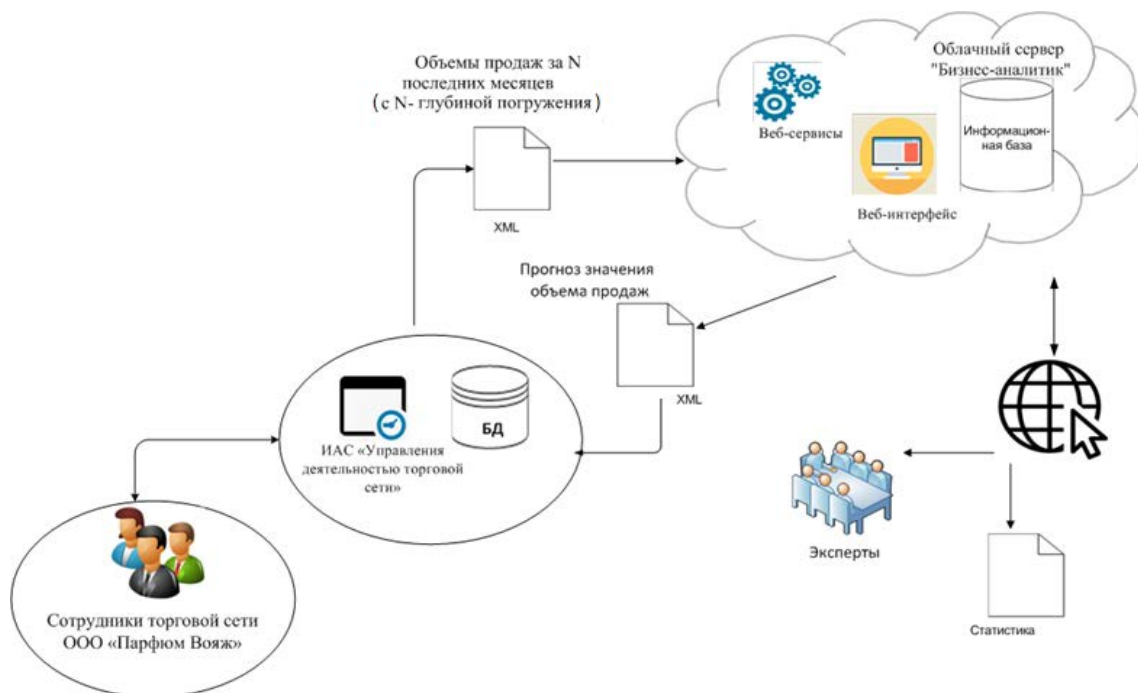


Рис. 4. Схема интеграции информационно-аналитической системы управления деятельностью торговой сети с WEB-сервисами системы «Бизнес Аналитик»

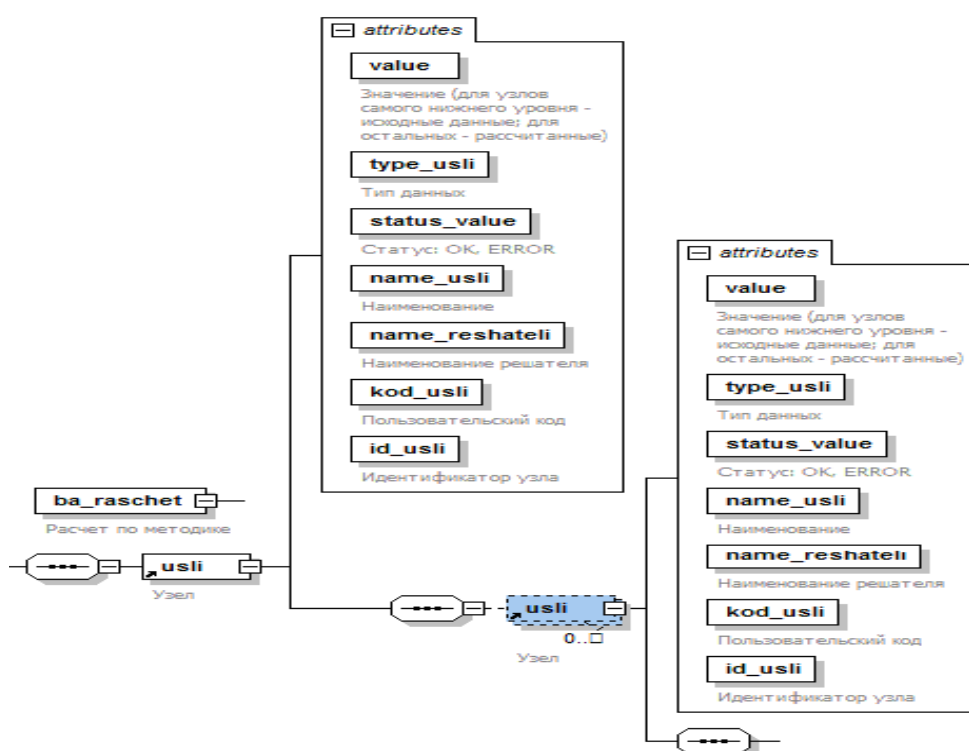


Рис. 5. Описание структуры XML файла – ответа WEB-службы при вызове метода «Расчет по методике»

Отчеты

Сервис

☆

☆

Интеграция СБА

Получить структуру данных БА

Выполнить прогноз

Еще

Продукция: Парфюмерная продукция (бизнес-класс)

N	Дата	Количество	Количество (прогноз)	Ошибка
54	01.06.2017	306	362	56
55	01.07.2017	390	538	148
56	01.08.2017	491	438	53
57	01.09.2017	363	408	45
58	01.10.2017	442	410	32
59	01.11.2017	405	371	34
60	01.12.2017	843	702	141
61	01.01.2018	241	270	29
62	01.02.2018	448	425	23

Рис. 6. Модуль интеграции с WEB-сервисами системы «Бизнес Аналитик» для автоматизации процедуры прогнозирования спроса на продукцию (объемов продаж) в составе информационно-аналитической системы управления деятельностью торговой сети

Модель оценки спроса на парфюмерную продукцию на основе факторного анализа (на примере туши).

Исходные данные: анкеты девушек в возрасте от 16 до 26 лет. Всего: 60 анкет.

Входные критерии: марка туши, щеточка, красота, стойкость, водостойкость, гипоаллергенность, цена (руб.). Выход: будете использовать вновь, будете рекомендовать подругам? (см. рис. 7).

Наименование	картинка	Почему выбрали именно эту туш	Рекомендовали или нет	Красота	Цена (руб.)
1. Maybelline LASH SENSATIONAL		1.удобная щеточка	Скорее да чем нет	Нравится	450
2. Maybelline LASH SENSATIONAL		1.Удобная щеточка 2.Стойкость	Однозначно да	Нравится	450
3. Loreal TELESCOPIC		стойкость щеточка	Однозначно да	Нравится	417
4. IsaDora Flashing Volume		Удобная щеточка	Однозначно да	Нравится	789

Рис. 7. Фрагмент исходных данных о тушах различных марок и отзывах девушек

Поскольку нейронные сети могут обучаться только на числовых данных, каждый качественный параметр был закодирован (см. таблицу 2).

Таблица 2

Таблица кодирования качественных факторов

	Показатель	Значение
Щеточка	удобная	1
	затрудняюсь ответить	0,5
	не удобная	0
Стойкость	стойкая	1
	затрудняюсь ответить	0,5
	не стойкая	0
Водостойкость	стойкая	1
	затрудняюсь ответить	0,5
	не стойкая	0
Гипоаллергенность	да	1
	затрудняюсь ответить	0,5
	нет	1
Красота	красивая	1
	затрудняюсь ответить	0,5
	не красивая	0
Выход (будете использовать вновь, будете рекомендовать друзьям?)	однозначно да	1
	скорее да, чем нет	0,5
	затрудняюсь ответить; скорее нет, чем да; нет	0

Цена туши закодирована через пропорцию. Область значений параметра «Цена» после кодирования: от 0 до 1 (0 – минимальная цена, 1 – максимальная цена).

Марки туши закодированы по принципу 1-из-N (см. рис. 8).

	A	B	C	D	E	F	G	H	I
1		Цена	Maybelline	Loreal	IsaDora	Pupa	MF	DIVAGE	Прочие
2	Maybelline	0,254843517	1	0	0	0	0	0	0
3	Loreal	0,230253353	0	1	0	0	0	0	0
4	IsaDora	0,507451565	0	0	1	0	0	0	0
5	PUPA	0,149031297	0	0	0	1	0	0	0
6	IsaDora	0,231743666	0	0	1	0	0	0	0

Рис. 8. Кодирование цены и марок косметической туши

Метод обучения нейронных сетей – kParTan.

Результаты вычислительных экспериментов показаны на рис. 9.

Наименьшая ошибка обобщения составила 0,153 при структуре 4-4-1.

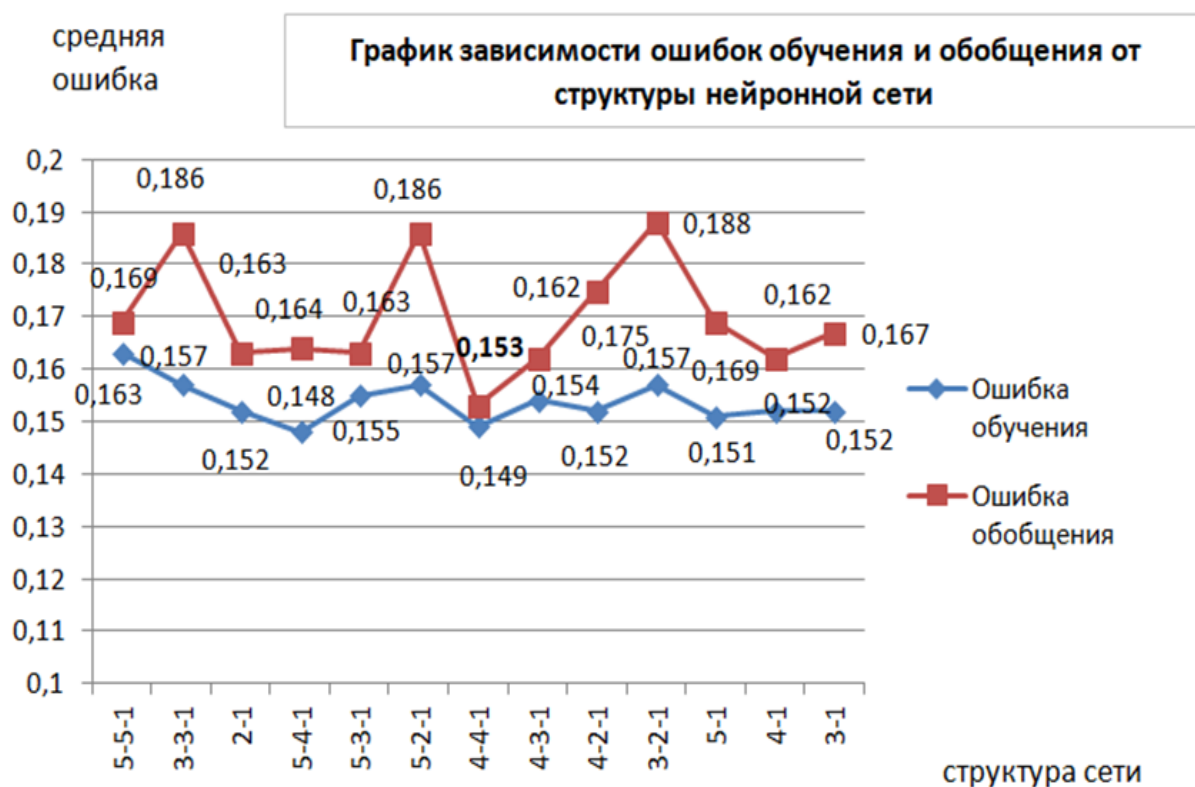


Рис. 9. Результаты вычислительных экспериментов. График зависимости ошибок обучения и обобщения от структуры нейронной сети

Полученные ошибки прогнозирования, ошибки обобщения свидетельствуют о достаточной адекватности построенных моделей оценки и прогноза спроса.

Безфакторная модель прогноза была апробирована на данных о суммарных объемах продаж (в штуках) парфюмерной продукции различных классов. В перспективе планируется по такой же технологии настроить модели прогноза спроса на конкретные виды продукции. Факторная модель оценки спроса на продукцию может быть использована с целью выдачи рекомендаций о принятии нового вида продукции на реализацию или нет.

В результате внедрения системы ожидается экономическая эффективность за счет повышения качества планирования закупок парфюмерной продукции у поставщиков и снижения объемов хранимой и списанной ввиду истечения срока годности «залежалой» продукции (по оценке экспертов примерно на 10%).

Список литературы

1. Горбань А.Н. Обучение нейронных сетей / А.Н. Горбань. - М.: изд. СССР-США СП "ParaGraph", 1990. - 159с.
2. Миркес Е.М. Нейрокомпьютер. Проект стандарта / Е.М. Миркес. - Новосибирск: Наука. Сибирское предприятие РАН, 1999. - 337с.
3. Попова Е.В., Гунер М.В., Пятковский О.И. Построение модели оценки и прогноза посещаемости предприятий общественного питания на основе применения нейросетевых технологий // Нейроинформатика, её приложения и анализ данных. Материалы XXV Всероссийского семинара. 2017. - с. 37-42.
4. Пятковский О.И., Гунер М.В., Авдеев А.С. Применение WEB-сервисов интеллектуальной информационной системы "Бизнес Аналитик" при решении задачи оценки инвестиционных проектов // Ползуновский альманах. 2013. № 1. - с. 102-105.

НЕЙРОСЕТЕВАЯ ОЦЕНКА ВЫБОРА МЕТОДА ЛЕЧЕНИЯ МКБ

А.В. Ершов¹, Ф.П. Капсаргин¹, К.В. Симонов²

¹ГБОУ ВПО КрасГМУ им. проф. В.Ф.Войно-Ясенецкого Минздрава России,
kapsargin@mail.ru

²Институт вычислительного моделирования СО РАН, *simonovkv50@gmail.com*

Введение

В настоящее время проявляется стремление медицины к объективизации путем количественного выражения клинических показателей. Причём это касается не только лабораторных данных, но и сведений, имеющих весьма субъективный характер оценки. Технологией, которая используется для решения подобных задач, являются искусственные нейронные сети.

Основное преимущество нейронных сетей в медицине – обработка и учет большого количества параметров, которые зачастую выявить врачу невозможно. В этом и заключается основная сложность прогнозирования, значительная часть информации представляет собой субъективные оценки врача, основанные на его знаниях и опыте [1, 2]. Тем не менее, влияние этих данных на постановку диагноза или выбора тактики лечения довольно высоко. Нейросети способны принимать решения, основываясь на скрытых закономерностях фильтруя многочисленные данные [3].

Нейросетевая модель кардинально отличается от других математических классификаторов. Во-первых, нейрокомпьютеры и нейропрограммы являются самообучающимися; во-вторых, принципы их работы очень напоминают (в сильно упрощенном виде) взаимодействие нейронов через синапсы [4, 5].

По доступным данным литературы, мы не обнаружили применения нейросетей в лечебном процессе при мочекаменной болезни (МКБ). Еще не разработаны прогностические экспертные систем, позволяющих выбирать наиболее приемлемый метод лечения в каждом конкретном случае, учитывая послеоперационные осложнения.

1. Исходные данные

Целью исследования являлась разработка экспертной системы распознавания и выбора метода лечения при МКБ на основе нейросетевого классификатора.

Обучающую выборку составили данные 510 пациентов (на базе урологических отделений КГБУЗ «Красноярская межрайонная клиническая больница скорой медицинской помощи им. Н.С. Карповича», НУЗ «Дорожная клиническая больница на ст. Красноярск ОАО «РЖД»).

Из поступивших больных 352 (69%) – мужчины, женщин – 158 (31%), из них 249 (48,8%) – поступили в плановом порядке, 261 (51,2%) – поступили по экстренным показаниям. Диагноз мочекаменной болезни впервые был установлен у 313 (61,4%), рецидивные конкременты выявлены у 197 (38,6%).

Анализ данных показал, что конкременты размерами от 0,8 до 2,5 см чаще локализовались в почечной лоханке 68 (13,4%) случаев, в чашечках – 46 (9,0%). Обструкцию пиело-уретерального сегмента камнем наблюдали у 16 пациентов.

Размеры конкрементов мочеточников колебались от 0,4 до 1,2 см, при локализации их: 174 (34,1%) в нижней трети, 132 (25,9%) в верхней трети, 90 (17,6%) в средней трети (табл. 1). Все пациенты обследованы по стандартному диагностическому алгоритму, включающему лабораторные и рентгенологические методы, которые позволяли выбрать оптимальный метод лечения в зависимости от анатомо-

функционального состояния верхних мочевых путей, локализации и размеров конкрементов, а также фазы течения пиелонефрита.

Уровень локализации	Пол		Оба пола %	Абс.
	Мужской пол %	Женский пол %		
Лоханка и чашечки почки	32,6	24,4	22,4	114
Верхняя треть мочеточника	13,0	16,4	25,9	132
Средняя треть мочеточника	24,8	14,2	17,6	90
Нижняя треть мочеточника	29,6	45,0	34,1	174
Левая сторона	40,2	67,6	44,0	224
Правая сторона	59,8	32,4	56,0	286

Дистанционная ударно-волновая литотрипсия проведена 141 (27,6%) пациенту, консервативная терапия проводилась в 125 (24,5%) случаях. Эндоурологические методы лечения (уретероскопия, перекутанная нефролитолапаксия) использовались в 110 (21,6%) и 81 (15,9%) случаях, соответственно. Остальные конкременты в удалены во время традиционных открытых операций – 53 (10,5%) случаев (рис. 1).

Уровень локализации	Пол		Оба пола %	Абс.
	Мужской пол %	Женский пол %		
Лоханка и чашечки почки	32,6	24,4	22,4	114
Верхняя треть мочеточника	13,0	16,4	25,9	132
Средняя треть мочеточника	24,8	14,2	17,6	90
Нижняя треть мочеточника	29,6	45,0	34,1	174
Левая сторона	40,2	67,6	44,0	224
Правая сторона	59,8	32,4	56,0	286

Таблица 1.



Рис. 1. Методы лечения больных с МКБ

2. Методика и результаты

Информация о каждом пациенте была представлена в виде многомерного вектора, характеризующаяся 28 входными параметрами: данные клинического осмотра, инструментального и лабораторного исследования.

Каждый пример имеет параметр «класс», представляющий заранее известную тактику лечения (консервативное лечение кодируется цифрой 0, уретероскопия КЛТ – цифрой 1, ДЛТ – 2, перкутанная нефролитолапаксия – цифрой 3, пиелолитотомия – 4, уретеролитотомия – 5). Количество примеров каждого класса составили все пациенты, прошедшие лечение в данных урологических отделениях.

В работе использовалась трехслойная нейронная сеть в среде Panalyzer 5.0 – наиболее оптимальная для решения задачи производства опыта из научных данных в гуманитарных областях. Обучение нейронной сети состоит в поиске закономерности и множества параметров, путем сравнения с эталонами. На выходе получаем определенный вывод, который делается на основании полученного опыта. Обучение сети для 28 входных векторов и 510 записей заняло около 65 минут.

Следующей важной задачей явилось изучение значимости входных параметров для постановки диагноза. Обученная нейросеть автоматически удалила наименее значимые из них. В результате, выделили 13 параметров, которые имели наибольшее влияние на выбор метода лечения (рис. 2).

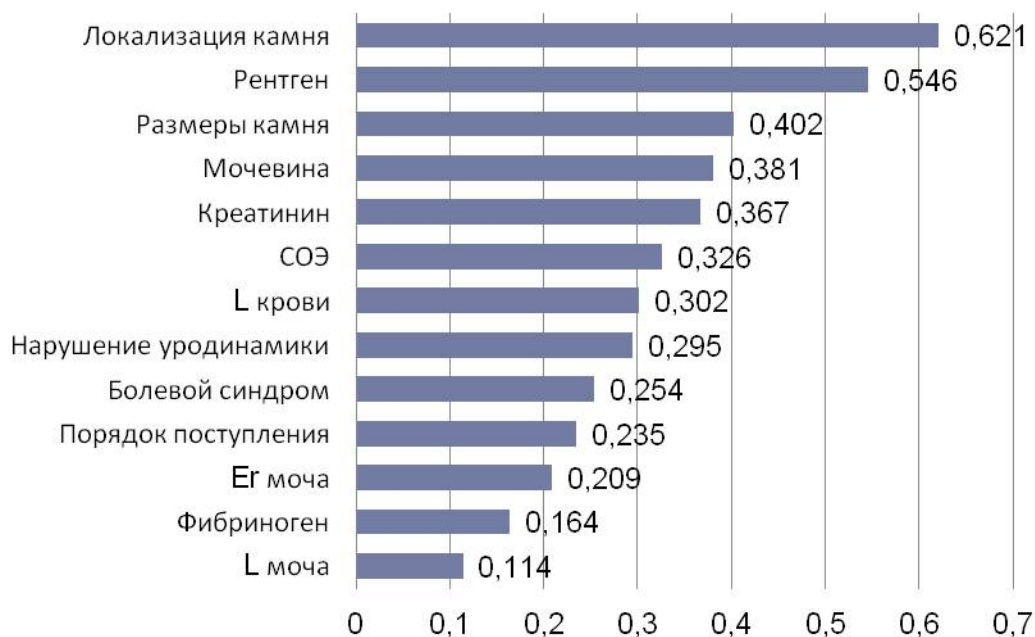


Рис. 2. Средняя значимость входных параметров

Наиболее значимыми показателями ($p < 0,05$) явились: локализация и размеры конкремента, видимость на снимках, показатели СОЭ, мочевины и креатинина крови, а также степень нарушения уродинамики.

Отличительной особенностью программы Panalyzer 5.0 является возможность создания отдельной программы – нейросетевой экспертной системы, позволяющей тестировать данные. Тестирование экспертной системы проводилось на 22 примерах. При этом нейронная сеть определила все предложенные примеры правильно. Более того, экспертная система «отбраковала» два примера, которые были внесены в тестовую выборку по ошибке. При ретроспективном анализе историй болезни данных пациентов выяснилось, что в обоих случаях конкременты нижней трети мочеточника самостоятельно отошли за стуки до проведения уретероскопии. Во время операции конкрементов в мочевых путях обнаружено не было.

Таким образом, решены следующие задачи:

- сформировано пространство информативных признаков, (в том числе данные субъективных оценок врача) влияющих на выбор метода лечения при мочекаменной болезни (МКБ);
- разработан нейросетевой классификатор в среде Panalyzer 5.0, рассчитаны показатели значимости параметров, которые имели наибольшее влияние на выбор метода лечения;
- выполнено тестовое прогнозирование разработанной экспертной системы с использованием контрольной выборки пациентов урологического отделения «Дорожной

клинической больницы на ст. Красноярск ОАО «РЖД», тестирование проводилось на 22 примерах, при этом степень диагностической уверенности системы составила 91%.

Заключение

Предложен нейросетевой метод выбора тактики лечения при мочекаменной болезни (МКБ). Обучение искусственной трёхслойной нейронной сети прямого распространения происходило по данным 510 историй болезни и многомерного вектора, характеризующегося 28 входными параметрами. Обучение нейронной сети состояло в поиске закономерности множества параметров, путем сравнения с эталонными результатами. Тестирование экспертной системы проводилось на 22 примерах. При этом нейронная сеть определила все предложенные примеры правильно. Разработана экспертная система, классифицирующая методы лечения (6 видов) МКБ со степенью уверенности в 91%.

Список литературы

1. Жилин В.В. Прогнозирование исхода процедуры неинвазивной элиминации конкрементов с использованием гибридных технологий нечеткой логики принятия решений и нейронных сетей: дисс. кан. тех. наук / В.В. Жилин. – Курск: б/и, 2009. – С. 125.
2. Круглов В. В. Искусственные нейронные сети. Теория и практика / В.В. Круглов, В.В. Борисов – М.: Горячая линия – Телеком, 2001. – С. 382.
3. Соломаха А.А. Современные тенденции прогнозирования в медицине / А.А. Соломаха, А.В. Костюнин, В.Г. Щетинин // Материалы 8-го Всероссийского семинара «Нейроинформатика и ее приложения» – Красноярск, 2000. – С. 162-163.
4. Горбань А.Н. Нейронные сети на персональном компьютере / А.Н. Горбань, Д.А. Россиев. – Новосибирск: Наука, 1996. – 276 с.
5. Россиев Д.А. Самообучающиеся нейросетевые экспертные системы в медицине: теория, методология, инструментарий, внедрение: дисс. док. мед. наук / Д.А. Россиев. – Красноярск: б/и, 1997. – С. 120.

РЕЗУЛЬТАТ ЭКСПЕРИМЕНТА ПО РАБОТЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ПОДБОРУ ПРИПУСКОВ ЗАГОТОВОК

А.И.Захарова

Сибирский федеральный университет,
Институт космических и информационных технологий, МБК «Прикладная физика и космические технологии, alex69_5@mail.ru

Для любого предприятия огромное значение имеет ритмичное функционирование, в процессе которого на каждом рабочем месте и участке производства, в каждом подразделении будет выполняться в данную единицу времени строго определенное количество работы и должностных обязанностей [1]. Данная проблема рассогласованности действий подразделений машиностроительного предприятия, влекущая за собой потери времени, простои цехов, была обнаружена в производстве АО «Информационные спутниковые системы» имени академика М.Ф. Решетнёва».

В АО «ИСС» действует процесс последовательного поэтапного изготовления детально-сборочных единиц (ДСЕ). Изначально, разработанная, согласованная и утвержденная конструкторская документация (КД) поступает в отдел планирования, где происходит просчет заготовки ДСЕ с нормативными, справочными припусками. Далее информация поступает на заготовительный участок, где происходит непосредственное нарезание заготовок, с последующей отправкой их в цех-изготовитель. Процесс не дает сбой в случае несложных, стандартных деталей, не требующих завышенных припусков на обработку, необходимых при изготовлении детали. Но, с появлением новых видов оборудования (многоцелевых станков с числовым программным управлением, обрабатывающих центров), новых инструментов и технологий изготовления, конструкции деталей начинают усложняться, технологические процессы приобретают специфику исполнения и отклонение от нормативного расхода металла [2]. Таким образом, возникает проблема согласования параметров заготовок ДСЕ между отделом планирования и технологами цехов-изготовителей.

Персонал отдела планирования имеет широкую специализацию работы, инженера не являются узконаправленными специалистами и не могут определить особенности изготовления той или иной сложной детали. Цеховой технолог, разрабатывая технологический процесс, может быть не удовлетворен размерами запланированной заготовки; возникает вопрос о пригодности заготовки для производства ДСЕ. При этом нельзя списать данный вид ошибки на человеческий фактор, ведь специалист отдела планирования верно рассчитал припуски по справочным нормативам, не учитывая пожелания цехового технолога, а технолог в свою очередь не имеет обязанности предоставлять четкое требование на каждую ДСЕ в условиях массового производства.

Для снижения частоты появления подобных ситуаций, персонал отдела планирования прибегает к согласованию размеров заготовок нестандартных деталей с цеховыми технологами. Возникает замкнутость системы, которая также может порождать простои заготовительного участка, когда специалист отдела планирования не готов в требуемый срок предоставить размеры заготовок, по причине ожидания ответа от технолога цеха. Возникают простои, потери времени в четком ритме производства.

При рассмотрении данной проблематики на других машиностроительных предприятиях отрасли, четкого решения проблемы выявить не удалось. На других производствах связь между подразделениями осуществляется также устным или письменным обоснованием размеров заготовок, базированием на прошлом опыте (неупорядоченные записи технологов отделов планирования), в ряде случаев используется метод консультации у сторонних специалистов. Ограниченность, или, наоборот, необоснованная сложность и дороговизна существующих решений проблемы очевидна. Ранее проводимые исследования по разработке эффективной и автоматизированной системы подбора параметров заготовок нами не выявлены.

Для решения данной проблемы разработана интеллектуальная система поддержки принятия решений (ИСППР), основанная на методе экспертных систем на прецедентах (casebasedreasoning, CBR). Это метод получения решений задачи путем поиска подобных проблемных ситуаций в памяти, хранящей прошлый опыт решения задач и адаптации найденных решений к новым условиям [3]. Основой системы является база знаний (БЗ), представляющая собой базу прецедентов, состоящую из 17 классов ДСЕ различных конфигураций, собранных в результате анализа деятельности отдела планирования АО «ИСС» за последние 4 года [4]. В ходе статистической обработки КД было рассмотрено 594 ДСЕ, выявлены основные параметрические характеристики, необходимые для разработки ИСППР (рис. 1).

Собранные данные были разделены на *априорные* (рис. 1, зеленый цвет), представленные специалистами отдела планирования, и *апостериорные* (рис. 1, розовый цвет), предоставленные технологами цехов-изготовителей (если такие данные имели место быть в виде корректировок размеров заготовок).

По итогам обработки каждого класса была определена генеральная совокупность прецедентов по классу, включающая *положительные* (без корректировки технолога цеха) и *отрицательные* (имеющие корректировку) прецеденты и выявлены производственные *правила* в количестве 44 штук, по которым наиболее верно происходит расчет размера заготовки в зависимости от ее формы, технологии изготовления и оборудования (участка в цехе-изготовителе).

№ п/п	Материал	Марка	Сортамент	ГОСТ/ТУ	Цех	Вес, кг	Чистовые размеры, мм			Раскрой, мм				Корректировка технолога, мм			
							a	b	c	Диаметр	Длина	Ширина	Высота	Диаметр	Длина	Ширина	Высота
1	Алюминий	АМГ6Б	Плита	ГОСТ17232-99	1	2,47	115	235	17,8	-	245	125	30	-	250(+5)	130(+5)	40(+10)
2	Алюминий	АМГ6М	Круг	ГОСТ21488-97	2	10,524	116	247	74	280	91	-	-	-	-	-	-
3	Титан	BT14	Плита	ОСТ190024-94	2	2,414	104,4	129,8	22	-	138	113	45	-	-	-	-
4	Титан	BT14	Плита	ОСТ190024-94	2	4,79	114,6	185,8	30	-	198	128	50	-	-	-	-
5	Титан	BT14	Плита	ОСТ190024-94	1	3,523	165	215	13	-	228	178	30	-	-	-	-
6	Титан	BT14	Плита	ОСТ190024-94	1	3,878	166	257	8	-	267	176	18	-	273(+6)	183(+7)	30(+12)
7	Титан	BT14	Плита	ОСТ190024-94	1	4,882	174	312	8	-	322	184	18	-	328(+6)	188(+4)	30(+12)
8	Титан	BT14	Плита	ОСТ190024-94	1	3,444	166	228	8	-	238	176	18	-	243(+5)	183(+7)	30(+12)
9	Титан	BT14	Плита	ОСТ190024-94	2	6,821	225	342,8	13	-	358	238	35	-	-	-	-
10	Титан	BT14	Плита	ОСТ190024-94	3	2,414	104,4	110	22	-	123	118	40	-	-	-	-
11	Титан	BT14	Плита	ОСТ190024-94	3	4,79	114,6	185,8	30	-	198	128	40	-	-	-	50(+10)
12	Титан	BT14	Плита	ОСТ190024-94	3	4,033	128	148	22	-	158	138	40	-	143(+5)	163(+5)	45(+5)
13	Титан	BT14	Плита	ОСТ190024-94	2	4,023	128	148	22	-	158	138	40	-	143(+5)	163(+5)	45(+5)
14	Алюминий	АМГ6М	Круг	ГОСТ21488-97	2	1,7	80	107	25	145	45	-	-	-	-	-	-
15	Алюминий	АМГ6М	Круг	ГОСТ21488-97	3	2,78	116	118	30	170	50	-	-	-	-	-	-
16	Алюминий	АМГ6М	Круг	ГОСТ21488-97	2	3,43	130	134	30	195	50	-	-	-	-	-	-
17	Алюминий	АМГ6М	Круг	ГОСТ21488-97	2	3,43	130	134	30	195	50	-	-	-	-	-	-
18	Алюминий	Д16ПП	Круг	ГОСТ21488-97	2	4,48	135	135	36,5	200	50	-	-	-	60(+10)	-	-
19	Алюминий	Д16ПП	Круг	ГОСТ21488-97	2	9,56	135	183,5	36,5	240	50	-	-	-	60(+10)	-	-
20	Алюминий	Д16Б	Плита	ГОСТ17232-99	3	5,846	135	156,5	40	-	172	150	60	-	-	-	-
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Рис. 1. Фрагмент набора данных, характеризующих 1-й класс ДСЕ [4]

На основании разработанного классификатора, системы правил, и алгоритма работы ИСППР была разработана программа, позволяющая делать выводы о размере припуска заготовки ДСЕ с использованием экспертной системы на прецедентах.

Приложение имеет вид, представленный на рис. 2. В левой части окна программы «Ввод данных» пользователь (работник отдела планирования) выбирает вручную априорные данные, характеризующие ДСЕ, согласно поступившей КД. В данном окне учтены все характеристики согласно параметрической модели заготовки для всех 17 классов деталей.

№	Диаметр, мм	Длина, мм	Ширина, мм	Высота, мм	Кэфф. уверенности	Значение EM
1	47	100	65	----	0.97	105.899
2	82	110	50	----	0.99	107.048
3	0	55	105	----	1.00	107.355
4	80	100	70	----	0.99	107.756
5	99	60	108	----	0.96	109.454

№	Материал детали	Марка материала	Сортамент заготовки	ГОСТ/ТУ	Дата изменения	ФИО	
1	80	Алюминий	АМГБМ	Круг	ГОСТ21488-97	22.10.2016	Кулаков Емельян Назарович

Рис. 2. Приложение по подбору припусков заготовок ДСЕ

Правое окно «Вывод результата» содержит совокупность прецедентов, имеющих схожие характеристики с расчетным, выводятся решения в виде размеров припусков для данных прецедентов. Цветом поясняется вид прецедента: зеленый – положительный (то есть не имеющий корректировок технологом цеха) или красный – отрицательный (имеющий корректировку). Также таблица содержит коэффициент уверенности, характеризующий степень адекватности выданных вариантов решения проблемы.

Нижняя таблица в окне приложения «Детализация прецедента» обеспечивает подробное раскрытие характеристик выбранного решения. Становится активной при нажатии на выданный ИСППР прецедент в правом окне программы.

Данное приложение было предложено сотрудникам АО «ИСС» им. ак. М.Ф. Решетнева для апробации в условиях производства.

Для проведения эксперимента были привлечены 5 работников отдела планирования, как лица, принимающие решения при расчете припусков заготовок ДСЕ, и один цеховой технолог, специфика производственной деятельности которого вынуждает его достаточно часто корректировать размеры припусков сложных ДСЕ для изготовления их на обрабатывающем центре модели Hermle C40U Dinamic.

Экспертам были предложены априорные данные 5 деталей сложной конфигурации, которые соответствуют различным классам классификатора ДСЕ,

разработанного в ходе диссертационного исследования. Априорные характеристики деталей представлены в табл. 1. На данные ДСЕ размеры припусков ранее не определялись, КД имеет первичный вид.

Таблица 1

Априорные характеристики экспериментальных ДСЕ

Параметр	1 деталь			2 деталь			3 деталь			4 деталь			5 деталь		
Класс детали	1			3			6			7			9		
Материал	Алюминий			Титан			Алюминий			Сталь			Алюминий		
Марка	Д16ТПП			ВТ14			АМГ6М			12Х18Н10Т			АМГ6М		
Сортамент	Круг			Плита			Круг			Круг			Круг		
ГОСТ/ТУ	ГОСТ 21488-97			ОСТ 190024-94			ГОСТ 21488-97			ТУ 14-1-3564-83			ГОСТ 21488-97		
Цех-изготовитель	1			1			2			3			2		
Чистовые размеры ДСЕ (ширина, высота, длина), мм	26	94	102	22	128	148	48	66	74	43	57	100	46	50	173

В табл. 2 представлены результаты проведенного эксперимента. Мнения экспертов в части размеров припусков заготовок для всех 5 ДСЕ были занесены в табл. 2. Следующим шагом априорные характеристики 5 экспериментальных ДСЕ были внесены в программу и произведен расчет размеров припусков с помощью ИСППР. Получившийся результат занесен в табл. 2 и выдан на оценку 6 экспертам. Эксперты №1, №2, №3, №4 и №5 – сотрудники отдела планирования, эксперт №6 – технолог цеха-изготовителя ДСЕ, специализирующийся на разработке технологии изготовления деталей сложных конфигураций. Перед экспертами стояла задача оценить адекватность принимаемого разработанным приложением решения согласно следующей шкале оценок [5]: 1 – очень плохо; 2 – плохо; 3 – удовлетворительно; 4 – хорошо; 5 – отлично. Получившиеся экспертные оценки занесены в табл. 2. Следует пояснить, что для 1 и 3 детали решение, принятое ИСППР, представляет собой средневзвешенную оценку параметров припуска группы положительных прецедентов. Для 2, 4 и 5 детали решение принималось ИСППР по системе разработанных правил для 3, 7 и 9 класса ДСЕ соответственно.

Анализ результатов апробации ИСППР в условиях реального производства заключается в анализе 2 показателей: коэффициента корреляции и средней экспертной оценки адекватности принимаемого ИСППР решения.

Коэффициент корреляции – это статистический показатель зависимости двух случайных величин. Коэффициент корреляции может принимать значения от -1 до +1. При этом значение -1 будет говорить об отсутствии корреляции между величинами, 0 – о нулевой корреляции, а +1 – о полной корреляции величин. То есть, чем ближе значение коэффициента корреляции к +1, тем сильнее связь между двумя случайными величинами [6].

Таблица 2

Экспертные оценки размеров заготовок ДСЕ

Мнение экспертов					Мнение ИСППР				Оценк а ¹
№ экспер та	Диамет р, мм	Длин а, мм	Шири на, мм	Высот а, мм	Диамет р, мм	Длин а, мм	Шири на, мм	Высот а, мм	
1 деталь									
1	100	110	-	-	105	110	-	-	4
2	105	110	-	-					5
3	150	35	-	-					4
4	105	110	-	-					5
5	105	115	-	-					5
6	105	110	-	-					5
2 деталь									
1	-	160	140	30	-	155	135	30	4
2	-	160	140	30					4
3	-	160	140	35					3
4	-	155	135	30					5
5	-	155	135	30					5
6	-	155	135	35					3
3 деталь									
1	90	85	-	-	85	100	-	-	3

¹ Оценка экспертами №1-6 адекватности принимаемого ИСППР решения.

2	85	80	-	-					5
3	90	85	-	-					3
4	90	90	-	-					4
5	90	100	-	-					4
6	90	100	-	-					4
4 деталь									
1	80	110	-	-	75	120	-	-	4
2	80	110	-	-					4
3	75	110	-	-					5
4	75	110	-	-					5
5	80	120	-	-					4
6	75	120	-	-					5
5 деталь									
1	75	185	-	-	75	200	-	-	4
2	75	185	-	-					4
3	75	195	-	-					5
4	75	190	-	-					4
5	75	200	-	-					5
6	75	200	-	-					5

Формула расчета коэффициента корреляции следующая [7]:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \times (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \times \sum (y_i - \bar{y})^2}},$$

где x_i – значения одной переменной из выборки, X ; y_i – значения другой переменной из выборки, Y ; $\bar{x}$ – среднее арифметическое для переменной X ; $\bar{y}$ – среднее арифметическое для переменной Y .

Средняя оценка рассчитывается стандартным способом. Оба показателя вначале рассчитываются для каждой детали, а потом находятся общие коэффициенты для экспериментальной выборки. Полученные данные представлены в таблице 3.

Результаты апробации технологии

№ детали	Коэффициент корреляции						Оценка адекватности принимаемого ИСППР решения	
	Диаметр	Длина	Ширина	Высота	Средний по детали	Средний по 5 деталям	Средняя по детали	Средняя по 5 деталям
1	0,173	-0,184	-	-	-0,006	0,451	4,7	4,3
2	-	0,447	0,447	-0,633	0,509		4,0	
3	-0,201	-0,585	-	-	-0,393		3,8	
4	0,447	-0,632	-	-	-0,093		4,5	
5	1	-0,533	-	-	0,233		4,5	

Таким образом, по результатам анализа показателей работы технологии интеллектуального принятия решений по подбору припусков заготовок, общий коэффициент корреляции на выборку пяти экспериментальных деталей $r = 0,451$, что говорит о высоком совпадении мнения экспертов и мнения ИСППР в части принятия решений; средняя экспертная оценка работы ИСППР – 4,3, что говорит о достаточно неплохом результате работы системы.

Список литературы

1. Туровец О.Г. Организация производства и управление предприятием / О.Г. Туровец, В.Б. Родионов // Экономика предприятия. — М.: Инфра-М. — 2004. — С.66 – 67.
2. Макареня Т.А. Организация и планирование машиностроительного производства / Т.А. Макареня // Машиностроение. — Таганрог: ТПИ. — 2007. — С.231 – 233.
3. Гаврилова Т.А. Базы знаний интеллектуальных систем / Т.А. Гаврилова, В.Ф. Хорошевский. — СПб.: Питер, 2001. — С.384.
4. Захарова А.И. Разработка базы знаний для интеллектуальной системы поддержки принятия решений по подбору припусков заготовок деталей в металлообрабатывающем производстве / А.И. Захарова, В.А. Углев // Робототехника и искусственный интеллект: Материалы IX Всероссийской научно-технической конференции с международным участием. — Красноярск: ЛИТЕРА-принт, 2017. — С.108-115.
5. Джексон П. Введение в экспертные системы / П. Джексон. — М.: Вильямс, 2001. — 624 с.
6. Елисеева И.И. Общая теория статистики / И.И. Елисеева, М.М. Юзбашев. — М.: Финансы и статистика, 2004. — 656 с.
7. Гусаров В.М. Статистика / В.М. Гусаров. — М.: Юнити-Дана, 2003. — 438 с.

РАЗРАБОТКА АИС СОДЕЙСТВИЯ ТРУДОУСТРОЙСТВУ ВЫПУСКНИКОВ С ФУНКЦИЕЙ ОЦЕНКИ ПРОФПРИГОДНОСТИ НА ОСНОВЕ ГИБРИДНОЙ ЭКСПЕРТНОЙ СИСТЕМЫ

Н.А. Звеков, Н.В. Смыкова, О.И. Пятковский

Алтайский государственный технический университет им. И.И. Ползунова,
кафедра информационных систем в экономике

Сейчас на рынке работает большое количество сервисов по трудоустройству, однако многие организации предпочитают напрямую обращаться в вузы и искать потенциальных сотрудников из числа выпускников. Для вуза становится актуальной задача автоматизации данного процесса с целью хранения информации о вакансиях и своевременном доведении ее до сведения выпускников. Поэтому некоторые университеты уже запустили собственные сервисы для содействия трудоустройству, например, Карьера. НГТУ, Flamingo (ТГУ) и другие. Однако помимо простой организации информационного взаимодействия между работодателями и выпускниками есть проблема профессионального отбора. Реализация подсистемы оценки профпригодности может увеличить долю эффективного трудоустройства, то есть когда на вакансию принимается соискатель в большей степени соответствующий предъявляемым требованиям.

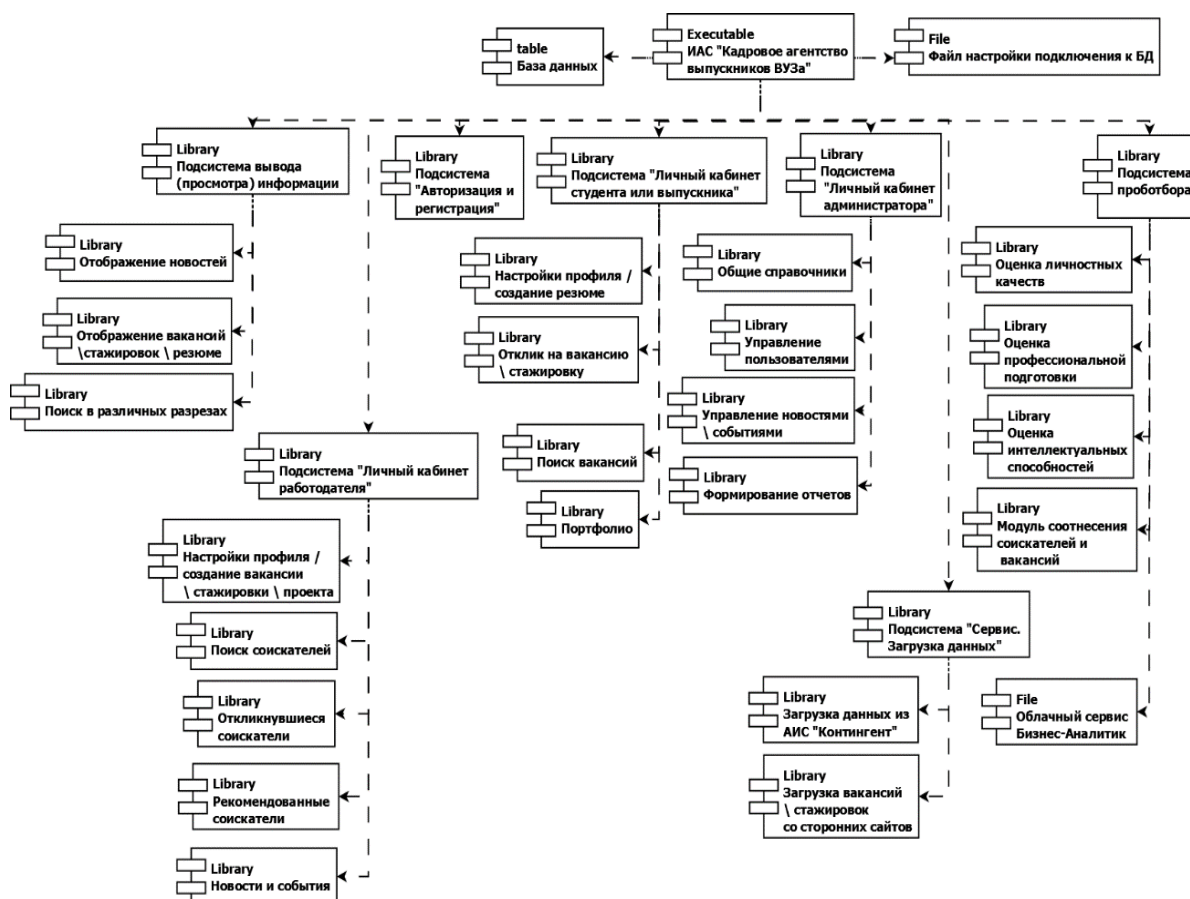


Рис. 1. Архитектура системы

Целью данной работы является разработка на основе web-технологий аналитической информационной системы трудоустройства выпускников вуза с

элементами оценки профпригодности выпускников на основе гибридной экспертной системы.

В качестве языка программирования для реализации системы был выбран PHP, СУБД MySQL. Предпочтение данным инструментам было отдано благодаря простоте и скорости разработки подобных систем. Архитектура системы включает подсистемы авторизации и регистрации, отображения новостей и поиска по различным параметрам, личный кабинет работодателя, личный кабинет студента или выпускника, личный кабинет администратора и подсистема профотбора (Рис. 1), в основе которой лежит гибридная экспертная система (Рис. 2).

В качестве решателей в гибридной экспертной системе использовались: формула, экспертная система и нейронная сеть. Эксперименты проводились на примере должности системного аналитика. Оценка профпригодности складывается из оценок личностных качеств, интеллектуальных способностей и профессиональной подготовки. Для получения первых двух оценок используются уже известные методики: краткий ориентировочный тест и 6-факторный личностный опросник Кеттелла соответственно. Оценка профессиональной подготовки основывается на требованиях профессиональных стандартов. Для каждой трудовой функции определен перечень дисциплин, в ходе изучения которых студент приобретает соответствующие знания и навыки, а исходными данными для оценки подготовленности к выполнению конкретной трудовой функции являются данные об успеваемости.

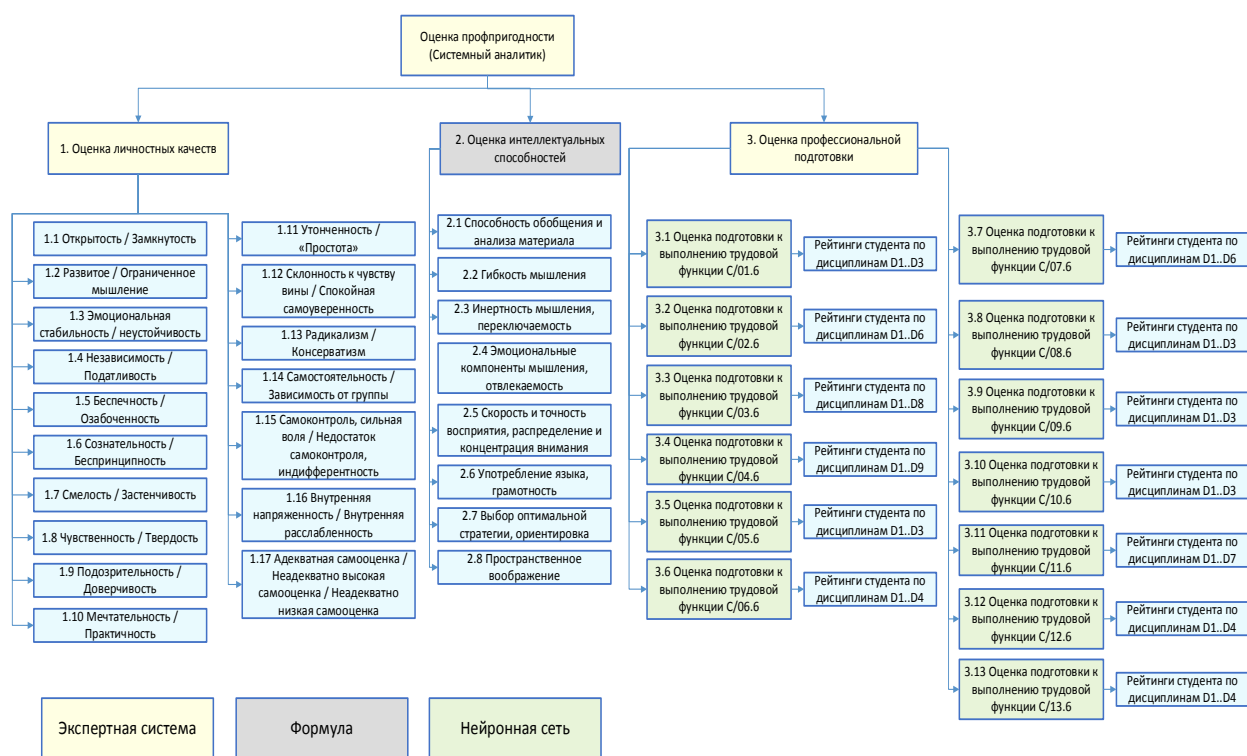


Рис.2. Иерархическая модель оценки профпригодности

Для работы с гибридной экспертной системой используется web-сервис Бизнес-Аналитик, в котором была настроена методика (Рис. 3). Между АИС и web-сервисом происходит обмен данными в формате xml (Рис. 4).

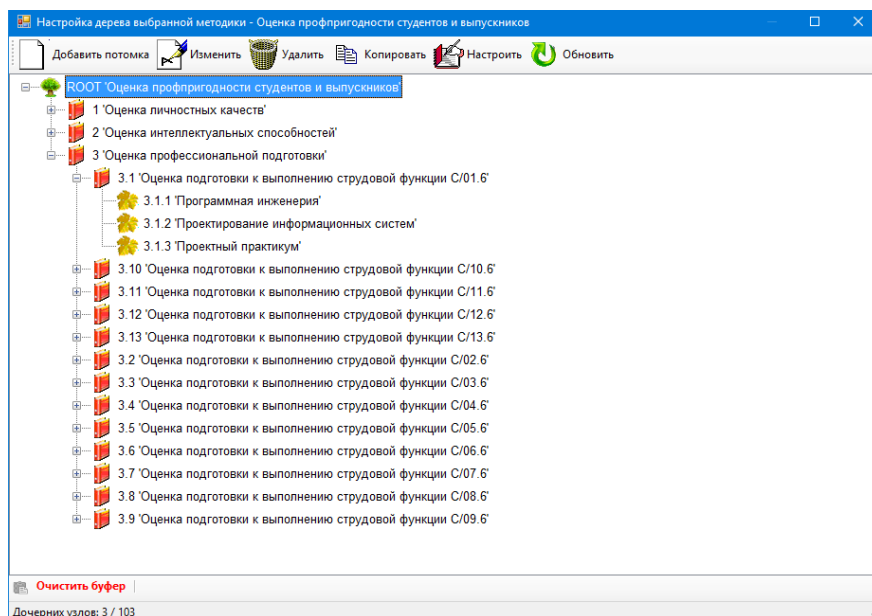


Рис. 3. Настройка методики в Бизнес-Аналитик

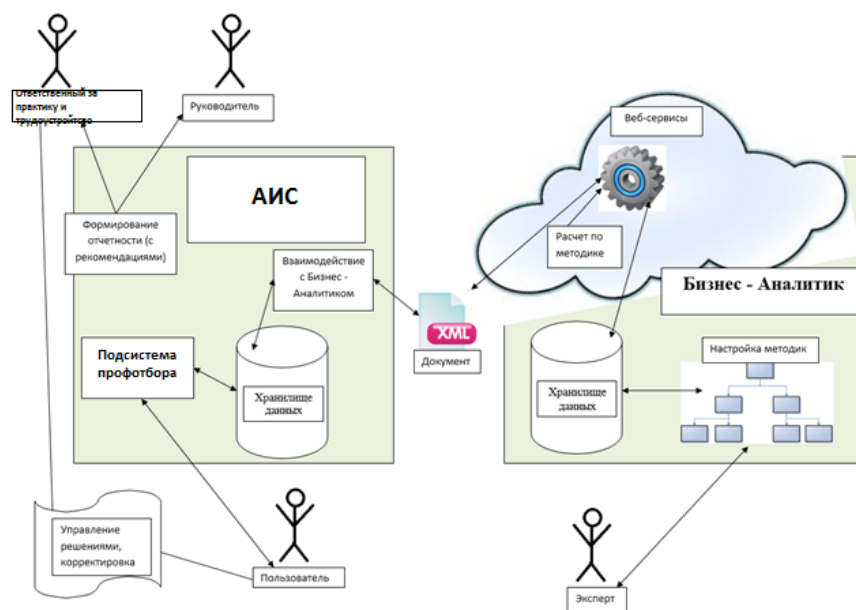


Рис. 4. Схема интеграции аналитической информационной системы

Разработанная методика была апробирована на данных выпускников текущего года. Пример расчета профпригодности приведен на рисунке 5. На основании полученных оценок профпригодности было произведено тестовое распределение выпускников (Рис. 6).

Для подтверждения достоверности полученных результатов предполагается сбор данных о фактическом трудоустройстве выпускников и дальнейшая обработка результатов.

Оценка профпригодности студентов и выпускников				
Рассчитать				
Код	Наименование	Значение		
1	Оценка профпригодности студентов и выпускников	3.4	Оценка подготовки к выполнению трудовой функции	4.6472
1.1	Открытость / Закрытость	3.1.1	Программная инженерия	55
1.10	Мечтательность / Практичность	3.1.12	1-ая производственная практика	75
1.11	Уточненность / "Простота"	3.1.13	Преддипломная практика	85
1.13	Радикализм / Консерватизм	3.1.14	Управление проектами	80
1.14	Самостоятельность / Зависимость от группы	3.1.2	Проектирование информационных систем	57
1.15	Самоконтроль, сильная воля / Недостаток самоконтроля, индифферентность	3.1.3	Проектный практикум	50
1.16	Внутренняя напряженность / Внутренняя расслабленность	3.1.6	Теория систем и системный анализ	75
1.17	Адекватная самооценка / Неадекватно высокая самооценка / Неадекватно низкая самооценка	3.1.7	2-ая производственная практика	75
1.2	Развитое / Ограниченное мышление	3.1.8	Базы данных	75
1.3	Эмоциональная стабильность / Неустойчивость	3.5	Оценка подготовки к выполнению трудовой функции	4.23417448173658
1.4	Независимость / Податливость	3.1.1	Программная инженерия	55
1.5	Беспечность / Озабоченность	3.1.2	Проектирование информационных систем	57
1.6	Сознательность / Беспринципность	3.1.3	Проектный практикум	50
1.7	Смелость / Застенчивость	3.6	Оценка подготовки к выполнению трудовой функции	3.5267
1.8	Чувствительность / Твердость	3.1.1	Программная инженерия	55
1.9	Подозрительность / Доверчивость	3.1.2	Проектирование информационных систем	57
2	Оценка интеллектуальных способностей	3.1.3	Проектный практикум	50
2.1	Способность обобщения и анализа материала	3.1.7	2-ая производственная практика	75
2.2	Гибкость мышления	3.7	Оценка подготовки к выполнению трудовой функции	4.14566097132465
2.3	Инертность мышления, переключаемость	3.1.10	Преддипломная практика	90
2.4	Эмоциональные компоненты мышления, отвлекаемость	3.1.2	Проектирование информационных систем	57
2.5	Скорость и точность восприятия, распределение и концентрация внимания	3.1.4	Информационные системы в организации	65
2.6	Употребление языка, грамотность	3.1.5	Сетевые технологии в экономике	45
2.7	Выбор оптимальной стратегии, ориентировка	3.1.8	Базы данных	75
2.8	Пространственное воображение	3.9	Оценка подготовки к выполнению трудовой функции	4.1839697227754
3	Оценка профессиональной подготовки	3.1.10	Преддипломная практика	90
3.1	Оценка подготовки к выполнению трудовой функции	3.1.2	Проектирование информационных систем	57
		3.1.3	Проектный практикум	50

Рис. 5. Расчет оценки профпригодности студентов и выпускников

Печать			
Группа ВПИЗ-61			
ФИО	Фактическое место работы	Основное	Место работы, должность
Альфа Михаил Алексеевич		Свеет, Программист	Добавить
Андреев Андрей Андреевич	Алтайский государственный технический университет им. И.И. Ползунова, Ижевск	Алтайский государственный технический университет им. И.И. Ползунова, Ижевск	Добавить
Бокос Константин Сергеевич	Алтайский государственный технический университет им. И.И. Ползунова, Доцент		Добавить
Ефанова Марина Алексеевна		Свеет, Веб-разработчик	ООО Холд, Программист 1С
Звеков Никита Александрович	АлпГТУ, ГК, Программист		АРЕАТ, Web-разработчик, программист
Иванюшина Анна Николаевна		Сибирск, Тестировщик	АРЕАТ Менеджер по работе с клиентами ООО Торговый дом АПОХИТ, Переводчик итальянского, английского языков
Калаев Ринат Тагатович			Добавить
Катушонко Татьяна Юрьевна	ЗАО "Флекс Софтваре Системс", Программист	Ритейл Сервис, Тестировщик ПО	Добавить
Купцова Татьяна Алексеевна	Марин-ра, Аналитик		Добавить
Лебедева Наталья Сергеевна	Businesscore, Программист		Добавить
Макаров Сергей Игоревич		1С-Галлекс, Младший консультант 1С	Деловой Барнаул, кадровое агентство, Системный администратор
Меленцова Елена Викторовна	Марин-ра, Аналитик	ПРОКС, Менеджер по продажам	ПРОКС, Бухгалтер
Сотоя Роман Александрович		ПРОКС, Специалист по работе с клиентами	Добавить
Труничева Анна Игоревна	ООО Фриматек, Руководитель проектов	Марин-Ра, Программист баз данных	Добавить
Тумилес Дмитрий Владимирович			Добавить
Ягайкина Ирина Геннадьевна		1С-Галлекс, Франчайзинг, Сервис-менеджер	Добавить

Рис.6. Результат распределения выпускников

Список литературы

1. Пятковский О.И., Смыкова Н.В. Применение экспертных оценок для определения связей трудовых функций профессиональных стандартов и дисциплин учебного плана // Ползуновский альманах № 2/2016 Виртуальные и интеллектуальные системы (Выпуск посвящен Седьмой всероссийской научно-практической конференции ВИС-2016, 23 ноября). – Изд-во АлпГТУ. – 2012. – С.131-133.
2. Пятковский О.И., Смыкова Н.В., Звеков Н.В. Оценка подготовки студентов в соответствии с профессиональными стандартами с использованием нейронных сетей Виртуальное моделирование, прототипирование и промышленный дизайн [Электронный ресурс]: материалы IV Международной научно-практической конференции : в 3 т. / под общ. ред. В. А. Немтинова ; ФГБОУ ВО «ТГТУ». – Тамбов : Изд-во ФГБОУ ВО «ТГТУ», 2017. – Вып. 4. (стр. 366-370)

ГИБРИДНЫЕ РЕШАЮЩИЕ МОДУЛИ С ВИРТУАЛЬНЫМИ ПОТОКАМИ В КЛАССИФИКАТОРАХ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ СЛОЖНЫХ СИСТЕМ

А.В. Киселев, Т.В. Петрова, О.В. Шаталова
Юго-Западный государственный университет,
кафедра биомедицинской инженерии, *SFelist@gmail.com*

При исследовании, как состояния живой системы, так и при прогнозировании ее перехода из одного состояния в другое, необходимо учитывать латентные переменные, которые не входят в исследуемый комплекс параметров системы, на основе которого строится пространство информативных признаков [1, 2]. Для учета этих переменных необходима системная надстройка, имеющая самоорганизующуюся структуру. Идея построения этой надстройки состоит в следующем. Если данные в системе носят поточный характер, сущность которого заключается в том, что каждый поток формируется множеством источников данных (рис. 1), сгруппированных по некоторому принципу, а затем сжатых, вплоть до скаляра, по некоторому формальному методу, то к имеющимся потокам может быть добавлен еще один поток - виртуальный поток, представленный своим выходом, то есть агрегированными данными, которые смоделированы тем или иным методом, например, Монте-Карло, или получены на основе того или иного закона распределения вероятностей.

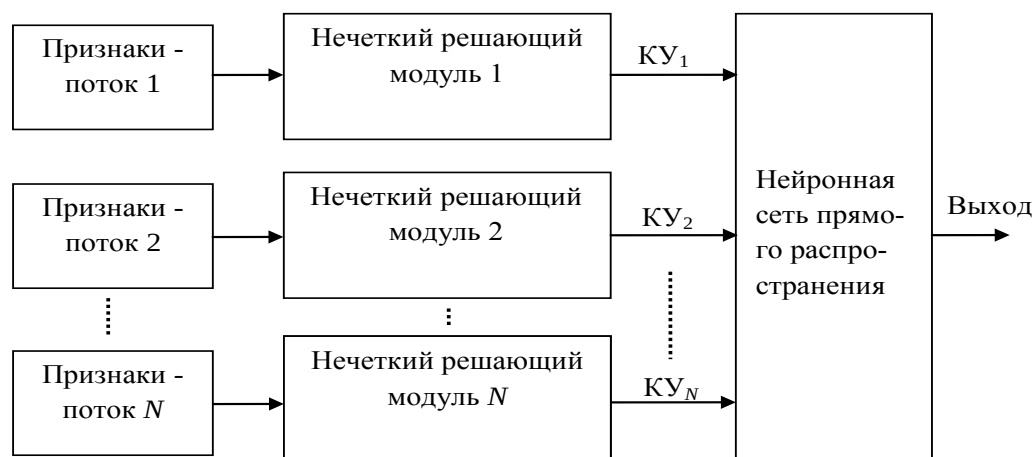


Рис. 1 - Структура гибридного решающего модуля с поточными данными

Если имеется функционал качества поведения системы, то меняя закон распределения вероятностей или его параметры в таком дополнительном потоке, можно добиться улучшения показателей качества системы или, хотя бы, Получив модель потока, можем генерировать данные в этом потоке, после чего их можно объединить с данными в других потоках, которые были получены посредством экспериментальных исследований на натуральных объектах (рис.1).

Для построения системной надстройки, реализующей работу с виртуальными потоками данных и интеграцию их в данные натурального эксперимента, необходим как

соответствующий инструментарий, так и методы и алгоритмы, позволяющие реализовать модель виртуального потока. В качестве инструментария для формирования виртуальных потоков используем гибридные решающие модули. Экспертные знания интегрируются в гибридный решающий модуль в виде нечеткого решающего модуля. На выходе нечеткого решающего модуля имеем скалярную величину, которую принято называть коэффициентом уверенности по данному сегменту информативных признаков или частным коэффициентом уверенности [3]. Агрегирование частных коэффициентов уверенности осуществляется на основе обучаемого классификатора, в качестве которого выступает нейронная сеть.

При использовании в качестве входов нейронной сети коэффициентов уверенности, получаемых в результате нечетких операций в нечетком решающем модуле, построенном на основе алгоритмов нечеткого вывода, возникают трудности с получением требуемой точности классификации нейронной сети, используемой на входе решающего модуля (рис. 1). Это связано с тем, что признаки, формируемые на входе нейронной сети, получены на основе экспертных заключений и гипотез, которые могут неадекватно отражать внутренние процессы в исследуемой системе. Кроме, того, полнота признакового пространства и релевантность самих признаков в нем определяется только экспертом (ЛПР). Эти признаки поступают на вход обучаемой нейронной сети, которая вносит свои искажения в интерпретацию признакового пространства. Особенно опасен здесь эффект переобучения, так как в случае прогнозирования заболеваний выборки достаточно большого объема сформировать весьма проблематично. Поэтому в структуру рис. 1 целесообразно внести дополнительные связи, которые бы позволили придать больший вес статистическим данным в общем процессе формирования гибридного решающего модуля.

Идея метода, позволяющего выполнить такую процедуру, состоит в следующем. Присутствующие на входе нейронной сети признаки -коэффициенты уверенности по сегментам информативных признаков - не отражают полностью свойства исследуемого объекта или системы, которые входят в область интересов ЛПР. Это утверждение справедливо хотя бы потому, что у ЛПР нет доказательств, что в N сегментах признаков на входе гибридного решающего модуля (рис. 1) собраны все релевантные признаки, характеризующие исследуемый объект, относящиеся к области интереса решаемой задачи или у ЛПР нет уверенности, что в гибридный решающий модуль включены все релевантные потоки.

Следовательно, можем выдвинуть гипотезу, что есть хотя бы один сегмент информативных признаков $N+1$, который включает хотя бы один информативный признак, коэффициент уверенности по которому (по сегменту $N+1$) KU_{N+1} , существенно влияет на качество принимаемых решений. Следовательно, необходима процедура, которая позволит сформировать вектор

$$\hat{Z}_{N+1} = (z_{N+1}^1, z_{N+1}^2, ..., z_{N+1}^i, ..., z_{N+1}^M)^T \quad (1)$$

для обучающей выборки

$$\{Z_1, Z_2, ..., Z_i, ..., Z_M\}, \quad (2)$$

где

$$Z_i = (z_1^i, z_2^i, \dots, z_j^i, \dots, z_N^i)^T, \quad (3)$$

M – число образцов в обучающей выборке, N – число входов нейронной сети на рис. 1, i – номер образца в обучающей выборке, j – номер входа агрегирующей нейронной сети ($j = \overline{1, N}$, $i = \overline{1, M}$).

Так как в (3) $z_j^i = KY_j^i$, то $Z_i = (KY_1^i, KY_2^i, \dots, KY_j^i, \dots, KY_N^i)^T$.

В общем случае у нейронной сети рис. 1 может быть введено неограниченное число дополнительных входов (имеется неограниченное число виртуальных информационных потоков), но в данном случае рассматривается только один дополнительный вход.

На поляризующий вход подаем константу 0,5; а на второй (дополнительный) – подаем случайную величину, распределенную по нормальному закону распределения, математическое ожидание которой равно 0,5. Получим зависимость количественных показателей качества классификации от дисперсии случайной величины на этом входе, принимая во внимание то, что используя величину дисперсии как «движок», мы имеем возможность, как включать, так и отключать этот вход.

«Включение» или «выключение» дополнительного входа осуществлялось посредством управления вариативностью данных, поступающих на него как в процессе обучения, так и при неизвестном образце. Если вариативность входа отсутствует, то он не оказывает влияние на принадлежность объекта к одному из двух альтернативных классов, а, следовательно, и на результаты классификации.

Качество классификации оцениваем через показатели чувствительности и специфичности [4]. Допустим, что количество объектов в тестовом наборе равно M , из них Mp – количество «положительных» объектов, а Mn – количество объектов «отрицательных», причем $M = Mp + Mn$. Пусть количество ложных пропусков FN , а ложных обнаружений FP , тогда можем подсчитать количество верных обнаружений TP и верных пропусков TN : $TP = Mp - FN$; $TN = Mn - FP$.

На контрольной выборке рассчитываются характеристики классификатора – чувствительность Se и специфичность Sp :

$$Se = \frac{TP}{TP + FN}, \quad (4)$$

$$Sp = \frac{TN}{TN + FP}. \quad (5)$$

Для количественного определения влияния дополнительного входа на качество классификации нейронной сети целесообразно два показателя (4) и (5) выразить в виде одного показателя. В качестве такого показателя часто используют диагностическую эффективность [4]. Однако у этого показателя имеется существенный недостаток, связанный с тем, что в результате простого объединения всех ошибок путем манипуляции соотношений положительных и отрицательных объектов в выборке этот показатель можно сделать достаточно большим. При этом классифицирующая система может быть полностью неработоспособна, например, иметь показатель (4) или (5)

равным нулю. Поэтому введем интегральный показатель качества классификации S , способ вычисления которого поясняет рис. 2.

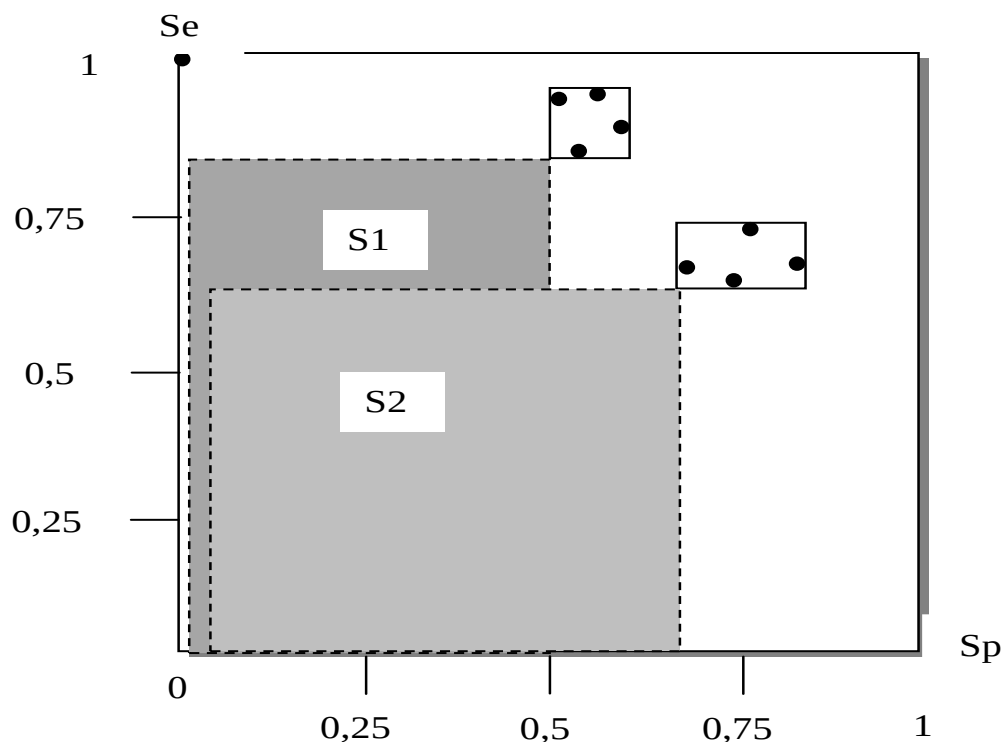


Рис. 2 – Определение интегрального показателя качества классификации

На рис. показатели качества классификации определяются кружками черного цвета. Кружки, вписанные в прямоугольник, это результаты классификации по выборкам из одной и той же генеральной совокупности. Для того, чтобы не загромождать рисунок, в прямоугольнике показаны только четыре кружка, которые определяют координаты и размер прямоугольника (лежат на нижней, верхней, правой и левой границах).

Показатель S определяет площадь заштрихованных прямоугольников, координаты которых определяют левые нижние углы прямоугольников, описывающие черные кружки. Для i -й выборки или группы выборок из одной и той же генеральной совокупности интегральный показатель качества классификации определяется по формуле

$$Si = \sqrt{Sei_{\min} \cdot Spi_{\min}}, \quad (6)$$

где $Sei_{\min}$ и $Spi_{\min}$ определяются координатами описывающих прямоугольников.

Экспериментально путем исследования искусственно сформированных выборок было установлено, что ввод («включение») дополнительного информативного признака (входа) в двухвходовой нейронной сети прямого распространения может повысить точность двухальтернативной классификации на 100% и выше, при разнице первых моментов у третьего признака всего лишь на 10%.

Для проведения этих экспериментальных исследований было создано программное обеспечение, позволяющее генерировать виртуальные классы в заданном

пространстве информативных признаков, и осуществлять визуализацию распределения классов в пространстве информативных признаков в двумерном или трехмерном пространствах. Для сравнительного анализа качества классификации формируемых нейросетевых структур использовался показатель (6). В процессе эксперимента генерировались обучающие выборки в трехмерном пространстве информативных признаков. Создавалась нейронная сеть для двухальтернативной классификации. Классы моделировались посредством управления математическим ожиданием и дисперсией генератора случайных чисел, распределенных по нормальному закону распределения.

Программное обеспечение формирования виртуального потока включает модуль генерации обучающей выборки и модуль генерации контрольной выборки. Модули генерируют массивы случайных чисел, распределенные по нормальному закону распределения, заданного размера. В качестве входной информации в этих модулях используются размерность пространства информативных признаков Dim , число образцов в обучающей или контрольной выборках M , математические ожидания компонентов векторов информативных признаков первого и второго классов, дисперсии компонентов векторов информативных признаков первого и второго классов $\{D_i\}$, $i=1,2...Dim$.

Модуль настройки нейронной сети позволяет формировать модель нейронной сети *NET* на основе обучающей выборки, сформированной модулем генерации обучающей выборки. В качестве входной информации для этого модуля, кроме обучающей выборки, необходимо задать число слоев нейронной сети и число нейронов в скрытых слоях. Необходимо учитывать, что минимальное число слоев нейронной сети – два, то есть мы можем исследовать нейронную сеть как минимум с одним скрытым слоем. В этом случае число нейронов в скрытом слое равно числу компонентов входного вектора и задается автоматически. В нашем эксперименте использовалась нейронная сеть с тремя слоями и с десятью нейронами в скрытых слоях. Нейронная сеть была построена в среде MATLAB [5].

Цель эксперимента заключается в том, чтобы определить, чувствителен ли вход третьего нейрона к вариации третьего информативного признака, если нейронную сеть настраивать при таких свойствах обучающей выборки. Иначе говоря, можем ли мы вводить дополнительные входы в нейронную сеть, подавая на них постоянный сигнал, и нейронная сеть при этом не обнулит веса соответствующего нейрона нулевого слоя. В случае наличия такой ситуации организовать обратную связь по дополнительному входу не представляется возможным, и все последующие теоретические построения станут неправомерными.

Настройка осуществлялась по сто элементной обучающей выборке, и нейронная сеть распределила по классам все образцы обучающей выборки правильно. Параметры нейронной сети, полученные при обучении, были сохранены в блоке *NET*.

Для проверки качества классификации нейронной сети *NET* генерируется контрольная выборка, компоненты векторов образцов которой имеют те же самые параметры распределения, что и обучающая выборка [6].

Следующим шагом эксперимента является изменение параметров закона распределения третьего информативного признака в очередной генерации контрольной

выборки: изменялась дисперсия третьего информативного признака в сторону увеличения. Дисперсия менялась от 10^{-6} до 10^{-2} . При изменении дисперсии на один порядок рост числа ошибок отражает следующий ряд: 0; 5; 32; 45; 50. Причем последние три числа являются, по существу, случайными числами.

Если дисперсию изменять у третьего признака только одного класса на те же величины, как и в первом случае, то получим следующий ряд ошибок: 1; 3; 13; 21; 26. При этом к случайной величине можно отнести только последний элемент этого ряда [6].

Для оценки степени управляемости показателями качества классификации нейронной сети посредством виртуального входа рассмотрим графики зависимости показателя качества классификации (S) от величины дисперсии виртуального входа, представленные на рис. 3. Управляемость величиной S посредством изменения дисперсии виртуального потока исследовалась при различных размерностях пространства информативных признаков 3, 5 и 7.

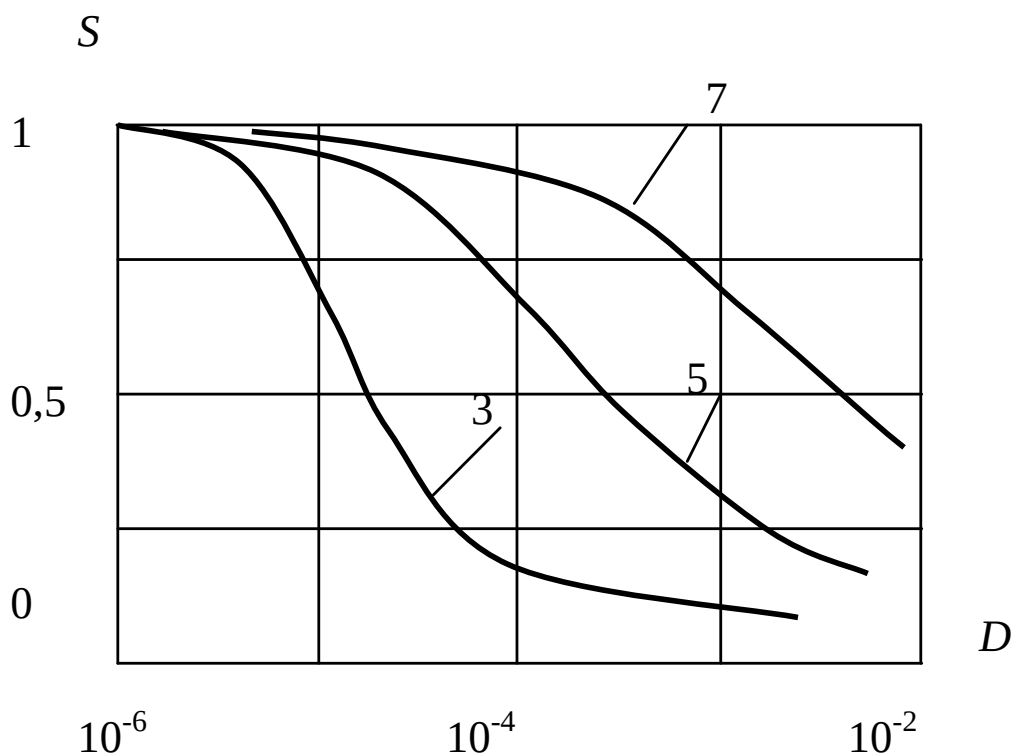


Рис. 3 – Влияние дисперсии виртуального входа на качество классификации нейронной сети прямого распространения при размерности пространства информативных признаков три, пять и семь

Представленные на рис. 3 зависимости показывают, что для широкого значения размерностей пространства информативных признаков может быть подобрана дисперсия виртуального потока, которая позволяет управлять показателем качества классификации нейронной сети в широком диапазоне. При этом управляемость показателя качества классификации снижается с ростом пространства информативных

признаков, откуда следует, что при увеличении размерности пространства информативных признаков необходимо увеличивать число виртуальных потоков.

Таким образом, введение виртуальных потоков в гибридный решающий модуль позволяет организовать «движок», осуществляющий управление качеством классификации нейросетевой структуры. С этой целью показатель качества классификации выбран таким образом, чтобы он зависел, с одной стороны, как от ошибок первого, так и второго рода, а с другой стороны, недопустимо большие ошибки хотя бы одного вида снижали бы показатель качества классификации к недопустимым значениям.

Список литературы

1. Позин, А.О. Синтез дополнительного информативного признака для нейронной сети прямого распространения / А.О. Позин, С.А. Филист, А.Н. Шуткин и др. // Новые решения в области упрочняющих технологий: взгляд молодых специалистов: сборник научных статей по материалам Международной научно-практической конференции. – Курск: ЗАО «Университетская книга», 2016. – С. 212 - 217.

2. Позин, А.О. Тестирующие гибридные системы с дополнительным пространством информативных признаков / А.О. Позин, С.А. Филист, А.Н. Шуткин // Современные информационные технологии в управлении качеством: сборник статей V Международной научно-прикладной конференции. – Пенза: Приволжский Дом знаний, 2016. – С.46-50.

3. Кореневский Н.А. Теория проектирования нечетких сетевых экспертных систем для управления медико-экологической безопасностью / Н.А. Кореневский, С.А. Филист, А.Б. Красковский и др. // Научные ведомости Белгородского государственного университета. Серия: История. Политология. Экономика. Информатика. 2009. № 9-1 (64). - Том 11. - С.146-151.

4. Филист С.А. Метод классификации сложных объектов на основе анализа структурных функций медленных волн / С.А. Филист, И.И. Волков, С.Г. Емельянов // Биомедицинская радиоэлектроника. №4. 2012. - С.6-11.

5. Леоненков, А.В. Нечеткое моделирование в среде Matlab и fuzzyTECH / А.В. Леоненков. – СПб.: БХВ – Петербург, 2005. – 736 с.

6. Шуткин, А.Н. Адаптируемые к структуре данных классификаторы в задачах прогнозирования профессиональных заболеваний / А.Н. Шуткин, А.О. Позин, С.А. Филист // Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сб. ст. по материалам Второй Всерос. науч.-практ. конф. – Пермь: Перм. гос. нац. исслед. ун-т, 2017. – С.161-164.

МЕТОДЫ ПРЕОБРАЗОВАНИЯ СИГНАЛОВ БОЛЬШОЙ РАЗМЕРНОСТИ В ДИАГНОСТИЧЕСКИХ МОДЕЛЯХ УПРАВЛЯЕМЫХ ДАННЫМИ

И.О.Лимарев¹, С.А.Субботин¹, А.А.Олейник¹, И.В.Дрокин²

¹Запорожский национальный технический университет,
Институт информатики и радиоэлектроники, каф. программных средств,
limarevigor@gmail.com, subbotin.csit@gmail.com, olejnika@gmail.com

²Центр испытаний УГК "Мотор Сич",
отдел исследований, *igor.dv.mail@gmail.com*

Преобразование сигналов является важной проблемой при построении аналитических и прогностических моделей, которые работают на основе интеллектуальных вычислений. Разработка таких моделей связана с необходимостью обработки больших данных [1].

Как правило, входные выборки данных которые представляют собой сигналы описывающие исследуемые процессы могут содержать избыточную и неинформативную информацию [2]. Использование такой информации при синтезе моделей приводит к повышению их сложности, избыточности, понижению обобщающих возможностей и может сделать синтезированную модель непригодной из-за необходимости использования больших вычислительных ресурсов.

Обычно, для решения этой проблемы проводят предварительную обработку данных для исключения из учебной выборки избыточных признаков и экземпляров. Большинство из известных методов используют жадную или стохастическую стратегию поиска. Обе эти стратегии не являются оптимальными для работы с временными последовательностями, которые представлены сигналами с большой частотой дискретизации из-за больших затрат временных и вычислительных ресурсов, а также сложности их реализации на практике.

Одним из эффективных инструментов анализа данных является цифровая обработка сигналов [3]. Она содержит большое количество инструментов для выделения необходимых компонентов сигнала, его нормализации и усреднения.

Рассмотрим задачу диагностики трансмиссии вертолета. Предотвращение отказу машинного оборудования является важным компонентом деятельности при обслуживании большинства инженерных систем. Вертолеты постоянно поддаются периодическим нагрузкам и вибрациям, которые инициируют возникновение повреждений в многих компонентах оборудования. Это связано с наличием в конструкции вертолета сложных механических систем: несущего и управляющего ротора, главной коробки передач и других элементов трансмиссии. В большинстве случаев, неисправности этих систем приводят к катастрофическим последствиям.

Для мониторинга технического состояния вертолета используются системы типа Health and Usage Monitoring Systems (HUMS) [4]. Эти системы позволяют обнаруживать повреждения в компонентах трансмиссии и прогнозировать их остаточный ресурс.

Важное значение в HUMS занимает возможность оценки технического состояния критически важных элементов трансмиссии, используя данные вибрационных сигналов, которые были записаны во время полета или наземных испытаний.

В данной работе ставится задача долгосрочного прогнозирования состояния трансмиссии вертолета во время наземного испытания. Такой тип испытания позволяет использовать значительные вычислительные мощности, что важно при работе с большим количеством сигналов, которые измеряются с большой частотой дискретизации и на значительных промежутках времени. Однако, при таком типе испытаний необходимо строить прогнозирующие модели, которые будут корректно работать на долгосрочных промежутках времени.

Для построения модели в решаемой задаче использовались данные вибродиагностики трансмиссии вертолета, которые содержат измерения пятнадцати датчиков, которые закреплены на компонентах трансмиссии, таких как: крышка вала ротора, стыковочный фланец верхней и нижней частей редуктора, фланец привода рулевого винта, промежуточный редуктор, хвостовой редуктор. В точке крепления располагаются 3 датчика, которые позволяют контролировать параметры вибраций в взаимно перпендикулярных плоскостях. Также измеряется сигнал оборотов двигателя.

Задача долгосрочного прогнозирования состояния трансмиссии вертолета включает в себя решение комплекса подзадач прогнозирования состояния отдельных элементов трансмиссии [5]. Обычно, эта задача решается путем прогнозирования следующих N значений определенного диагностирующего сигнала.

Однако, этот метод имеет недостатки, которые делают его сложным для применения на практике. Во-первых, входной сигнал имеет низкую информативность относительно состояния конкретного элемента трансмиссии. Во-вторых, состояние элемента трансмиссии обычно характеризуется не одним, а несколькими диагностирующими сигналами. В-третьих, входные сигналы имеют большую частоту дискретизации.

В то же время, есть потребность в построении прогнозирующих моделей на основе больших данных для обеспечения удовлетворимой точности прогноза. Эти ограничения делают процесс синтеза и обучения моделей на основе вычислительного интеллекта такими что требуют больших затрат временных и вычислительных ресурсов.

Для решения этих проблем было разработано два метода преобразования входных сигналов, которые базируются на методах цифровой обработки сигналов и могут быть применены для мониторинга и прогнозирования состояния определенного элемента трансмиссии при конкретизации параметров обработки сигнала.

Метод преобразования сигнала для получения значений по параметру виброперегрузки предлагается использовать для получения оценки состояния элемента трансмиссии по параметру виброперегрузки.

Этап инициализации. Задать входной сигнал $x(t)$.

Этап рекурсивной фильтрации. Применить к входному сигналу $x(t)$ фильтр Чебышева 1 рода для выделения из него необходимой полосы частот, которая характеризует состояние элемента трансмиссии. АЧХ (амплитудно – частотная характеристика) фильтра Чебышева 1 рода определяется в как

$$K(\omega) = \frac{1}{\sqrt{1 + \varepsilon^2 T_n^2(\omega / \omega_0)}}, \quad (1)$$

где ω_0 – частота среза, $T_n(x)$ – полином Чебышева n -го порядка, n – порядок фильтра, ε – параметр, который определяет величину пульсаций АЧХ в полосе пропускания.

Этап выделения огибающей сигнала. Выделить огибающую сигнала для получения абсолютного значения сигнала. Для этого, согласно с, представим полученный в результате предыдущего этапа сигнал в виде аналитического сигнала путем добавления к действительному сигналу $x_f(t)$ его преобразования Гильберта в качестве мнимой части:

$$\psi(t) = x_f(t) + j\hat{x}_f(t), \quad (2)$$

отсюда огибающая сигнала:

$$\omega(t) = |\psi(t)| = \sqrt{x_f^2(t) + \hat{x}_f^2(t)}. \quad (3)$$

Выделение огибающей сигнала производится с использованием пикового детектора, который служит для выделения как максимального, так и минимального значений огибающей сигнала.

Этап усреднения сигнала. Получить среднее значение амплитуды сигнала за τ секунд по формуле:

$$x'(p) = \frac{\sum_{i=1}^{H\tau} \omega(i)}{H\tau}, \quad p = 1, 2, \dots, P, \quad (4)$$

где H – частота дискретизации огибающей сигнала $\omega(t)$.

Метод преобразования сигнала для получения значений по параметру виброскорости предлагается использовать для получения оценки состояния элемента трансмиссии по параметру виброскорости.

Этап инициализации. Задать входной сигнал $x(t)$.

Этап рекурсивной фильтрации. Применить к входному сигналу $x(t)$ фильтр Чебышева 1 рода для выделения из него необходимой полосы частот, которая характеризует состояние элемента трансмиссии.

Этап интегрирования сигнала. Провести интегрирование сигнала методом RC-цепочки с числом точек усреднения N для перехода к параметру виброскорости. RC-цепь представляет собой электрическую цепь, которая состоит из резистора R и конденсатора C . Интегрирование происходит по формуле:

$$x_{\text{int}}(t) \approx \frac{1}{RC} \int_0^t x_f(t) dt, \quad x_{\text{int}}(t) \ll x_f(t), \quad (5)$$

где R – сопротивление резистора, C – емкость конденсатора, $x_f(t)$ – сигнал полученный в результате предыдущего этапа.

Этап выделения огибающей сигнала. Выделить огибающую сигнала (2)-(3) для получения абсолютного значения сигнала. Выделение огибающей сигнала производится с использованием пикового детектора, который служит для выделения как максимального, так и минимального значений огибающей сигнала.

Этап усреднения сигнала. Получить среднее значение амплитуды сигнала за τ секунд по формуле (5).

Разработанные методы преобразования сигналов позволяют выделить необходимые компоненты из исходных сигналов. После выделения необходимой компоненты проводится усреднение сигнала.

При помощи полосового фильтра происходит очищение сигналов от низкочастотных и высокочастотных компонент, которые являются малоинформативными в исходных сигналах. Дальнейшее усреднение сигнала позволяет понизить влияние случайных и шумовых составляющих сигнала, а также позволяет значительно понизить его размерность.

Это позволяет использовать обработанные сигналы в нейросетевых моделях более эффективно чем исходные. На практике это позволит увеличить скорость построения прогнозирующих моделей.

После обработки исходных сигналов на их основе можно построить прогнозирующую модель.

В большинстве HUMS в качестве прогнозирующих моделей применяются многошаровые нейронные сети прямого распространения сигнала. Эти модели демонстрируют свою работоспособность для кратковременного прогнозирования, но с увеличением временного окна между входными данными и целевым значением прогноза качество работы моделей падает.

Модели рекуррентных нейронных сетей имеют связи между элементами, которые создают направленную последовательность. Благодаря этому появляется возможность обрабатывать последовательные серии событий во времени. В отличие от многошаровых персептронов, рекуррентные нейронные сети могут использовать свою внутреннюю память для обработки последовательностей произвольной длины.

В данное время существует большое количество разновидностей моделей рекуррентных нейронных сетей для решения разнообразных задач. В [6] проведено анализ основных моделей рекуррентных нейронных сетей для решения задачи долгосрочного прогнозирования состояния трансмиссии вертолета.

На основе проведенных исследований, в качестве оптимальной нейросетевой модели была выбрана нейросеть нелинейной авторегрессии с внешними входами (Nonlinear Autoregression with eXternal inputs, NARX). Она является динамической нейронной сетью, которая имеет линии задержки на входе и глобальные рекуррентные связи.

Нейросети NARX являются более мощными рекуррентными нейросетями, чем простые рекуррентные нейросети (Simple Recurrent Network, SRN), поскольку имеют и рекуррентные связи, и линии задержки.

Для исследования предложенных методов преобразования сигналов, они были реализованы программно, используя процедуры обработки данных WinPOS. Нейросетевая модель была реализована на Python, при помощи библиотеки PyTorch.

В ходе проведения экспериментов из 16 исходных сигналов было выделено и получено 22 искусственных (рассчитанных) сигналов, которые отображают состояние таких элементов трансмиссии, как: зубозацепление главного редуктора, лопадка несущего винта, хвостовой вал, зубозацепление промежуточного редуктора, зубозацепление хвостового редуктора, лопадка рулевого винта, рулевой винт.

Для рекурсивной фильтрации в предложенных методах преобразования сигналов использовался фильтр 4 порядка с величиной пульсаций $\varepsilon = 0,3$ дБ. Для интегрирования сигнала методом RC-цепочки число точек усреднения $N = 800$. Сигналы усреднялись за время $\tau = 3$ с.

Полоса частот определялась отдельно для каждого элемента трансмиссии, соответственно его физическим параметрам. Например, для обработки сигнала по параметру хвостового вала использовалась полоса частот пропускания фильтра 20 – 45 Гц (Рис.).

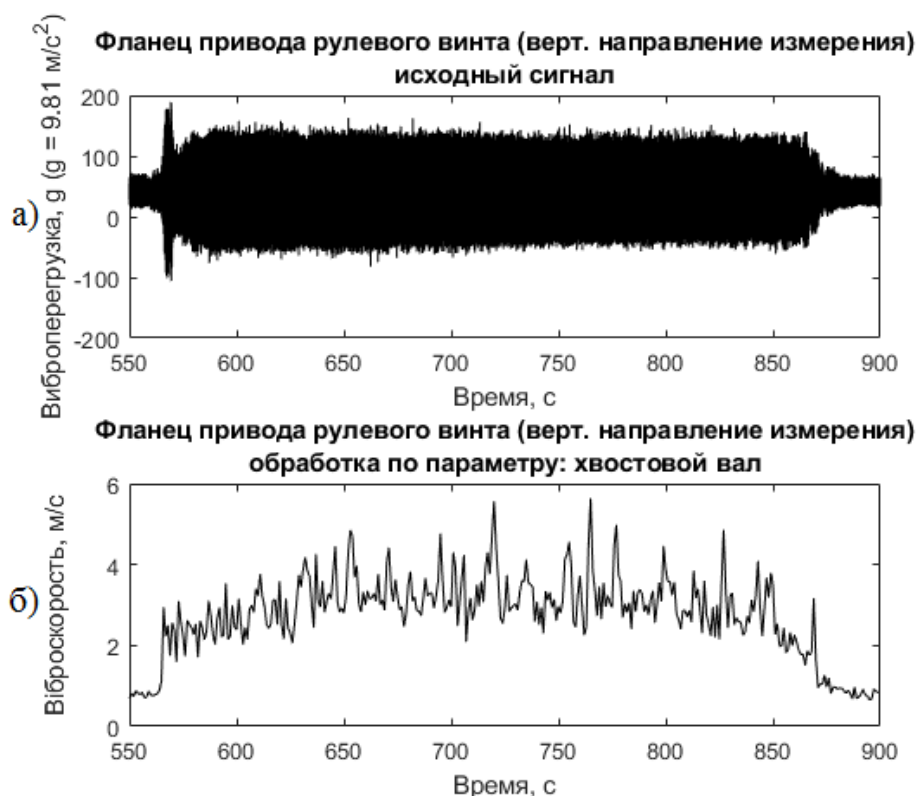


Рис. Исходный сигнал измерения вибраций для фланца привода рулевого винта (вертикальное направление измерения) (а) и сигнал после обработки предложенным методом по параметру – хвостовой вал (б)

Для проведения экспериментов использовалась нейросетевая модель NARX с параметрами: порядок линий задержки для входных значений $N = 2$, порядок линий задержек для исходящих значений $L = 10$, порядок линий задержек для рекуррентных связей $M = 5$.

Были проведены эксперименты, во время которых в качестве обучающей выборки использовалась историческая информация о ходе 5 циклов испытаний трансмиссии (длительность цикла $t = 150$ с). Исходные сигналы были преобразованы с использованием предложенных методов в сигналы, которые характеризуют состояние отдельных элементов трансмиссии.

Ставилась задача прогнозирования амплитудных значений этих сигналов во время следующего испытательного цикла. В качестве целевых сигналов, во время эксперимента, поочередно использовались все полученные сигналы. Для сравнения, по такой самой схеме эксперимента проводилось прогнозирование части входных необработанных сигналов, на основе исторической информации о значении всех остальных входных сигналов. Для оценки качества прогноза использовалась оценка средней точности прогноза по формуле:

$$Accuracy = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{N}, \quad (14)$$

где N – количество точек измерения сигнала, y_i – истинное значение амплитуды i -того отсчета сигнала, $\hat{y}_i$ – спрогнозированное значение амплитуды i -того отсчета сигнала.

В таблице приведён фрагмент полученных результатов экспериментов.

Таблица

Средняя точность прогноза значений сигналов за один испытательных цикл

Тип сигнала	Место и направление измерения	Параметр обработки	Средняя точность
Исходный необработанный	Фланец привода рулевого винта (верт. направление измерения)	-	0.912
	Промежуточный редуктор (гориз. направление измерения)		0.956
	Хвостовой редуктор (осевое направление измерения)		0.924
	Фланец привода рулевого винта (верт. направление измерения)	Хвостовой вал	0.912

Обработанный предложенными методами	Промежуточный редуктор (гориз. направление измерения)	Зубозацепление промежуточного редуктора	0.956
	Хвостовой редуктор (осевое направление измерения)	Лопатка рулевого винта	0.924

Результаты проведенных экспериментов показали, что средняя точность прогноза на основании необработанных сигналов и сигналов, которые были преобразованы с использованием предложенных методов, существенно не отличаются. Однако, в первом случае, обучение модели потребовало больших вычислительных мощностей. Использование преобразованных сигналов требует намного меньших мощностей, что облегчает практическое применение таких моделей, в том числе открывает возможности для проведения их обучения в онлайн-режиме, что является важным при решении задач технической диагностики.

Список литературы

1. Bishop C.M. Pattern recognition and machine learning / C.M. Bishop — Springer, New York, 2006. — 738 p.
2. Subbotin S.A. The dimensionality reduction methods based on computational intelligence in problems of object classification and diagnosis / S.A. Subbotin, A.A. Oliinyk // Advances in Intelligent Systems and Computing. — 2017. — Vol. 543. — P.11 – 19.
3. Ukil A. Intelligent systems and signal processing in power engineering / A. Ukil — Springer, Berlin, 2007. — 372 p.
4. Dempsey P. Investigation of Current Methods to Identify Helicopter Gear Health / P. Dempsey // 7 IEEE Aerospace Conference. — 2007. — P.1 – 14.
5. Kazlas P. T. Neural network-based helicopter gearbox health monitoring system / P. T. Kazlas, P. T. Monsen, M. J. LeBlanc // Neural Networks for Signal Processing, Proceedings of the IEEE-SP Workshop. — 1993. — P. 431 – 440.
6. Лимарев І.О. Довготермінове прогнозування стану трансмісії вертольоту на основі рекурентних нейронних мереж / І.О. Лимарев // Радіоелектроніка і молодь в ХХІ ст.: матеріали ХХІІ Міжнародного молодіжного форуму. — 2018. — Т. 5. — С. 223-224.

МОДЕЛИРОВАНИЕ ДИВЕРГЕНТНОГО МЫШЛЕНИЯ В МАЛЫХ КОЛЛЕКТИВАХ ЭКСПЕРТОВ ГИБРИДНЫМИ ИНТЕЛЛЕКТУАЛЬНЫМИ МНОГОАГЕНТНЫМИ СИСТЕМАМИ²

С.В. Листопад

Калининградский филиал Федерального исследовательского центра
«Информатика и управление» Российской академии наук, ser-list-post@yandex.ru

В связи с тем что применение абстрактно-математических моделей и традиционных аналитических подходов к принятию решений в динамической среде ограничено, актуально моделирование процессов решения проблем в горизонтальных структурах, позволяющих интегрировать разнородные знания и учитывать динамический характер среды принятия решений, таких как «круглый стол», «консилиум», «планерка» и др. Концепция интеграции разнородных, дополняющих друг друга знаний в интересах компьютерной поддержки индивидуальных решений воплотилась в гибридных интеллектуальных системах (ГиИС) А.В. Колесникова [1]. Недостаток ГиИС – наличие зависимости успешности решения проблемы от качества заменяющей ее системы задач и функционально соотнесенного с ними поля-множества моделей, над которым синтезируется интегрированный метод. Для преодоления этого недостатка предлагается разработка нового класса интеллектуальных систем – гибридных интеллектуальных многоагентных систем гетерогенного мышления (ГИМСГМ). ГИМСГМ сочетают аппарат многоагентных систем в смысле В.Б. Тарасова [2], и методики гетерогенного мышления: дивергентное, расходящееся, безоценочное восприятие проблемной ситуации и конвергентное, сходящееся мышление С. Кейнера [3], параллельного мышление, конструирование решений из поля параллельно существующих возможностей Э. де Боно [4] для развития и компьютерного моделирования и поддержки межличностного интеллекта в смысле Г. Гарднера [5] и динамики групповой работы сотрудничающих агентов, что релевантно воссоздает обслуживающую задачу в контексте проблемной ситуации с недостаточно выявленными или обнаруженными условиями и с неопределенной целью. В настоящей работе рассматриваются вопросы компьютерного моделирования дивергентного мышления.

Современный психологический анализ процессов решения задач и проблем дан В.Ф. Спиридоновым, утверждающего: решение задач и проблем протекает в форме процесса; мышление происходит в форме решения задач и проблем; мышление – процесс решения, разворачиваемый в реальном времени; мышление может быть целиком «внутренним», т.е. протекать безо всякой опоры на предметы или явления реального (внешнего) мира; мышление невидимо для своего носителя, мыслительным процессом трудно управлять; гарантированное успешное решение нестандартных задач или проблем – утопия [6].

Организация решений проблем малым коллективом экспертов заключается не только в том, чтобы собрать людей вместе и обозначить проблему. При решении вновь возникающих, ранее не встречавшихся задач, процесс, в общем случае, состоит из следующих этапов: постановка, формулирование и анализ задачи, сбор и интерпретация данных, поиск решений, анализ эффективности решений и окончательный выбор, представление результатов, реализация решения, мониторинг и оценка результатов [7].

² Работа выполнена при поддержке РФФИ (проект № 18-07-00448).

Периоды коллективного обсуждения общей задачи могут чередоваться с периодами индивидуальной работы экспертов над «своими» подзадачами. Процесс решения задачи накладывается на процесс формирования и развития коллектива экспертов как единой сущности, состоящий из стадий: формирование, бурление, становление норм, выполнение, принятие решения и расформирование [8, 9].

На первой стадии формирования члены коллектива знакомятся, обмениваются официальной информацией друг о друге, вносят предложения о работе коллектива, придерживаются общепринятых точек зрения, высказывают предложения, ведущие к очевидным решениям [9]. Если задача имеет очевидное решение, то дискуссия завершается уже на этом этапе. Проблема в том, что часто даже в сложных проблемных ситуациях дискуссия завершается еще на стадии формирования без выработки удовлетворительного решения. Основная ее причина – неспособность членов коллектива к совместной работе: различия в точках зрения рассматриваются как конфликты-барьеры и их преодоление, в результате перспективные нестандартные решения отбрасываются [3].

Если коллективу экспертов удалось выйти за границы устоявшихся мнений, процесс обсуждения переходит в стадию бурления, когда отдельные участники начинают высказывать нестандартные, часто противоречивые решения. И здесь возникают конфликты и конфронтация между членами коллектива. Если различия слишком велики, то некоторые члены могут покинуть коллектив, иначе они приспособляются или открыто обсуждают противоречия. Важную роль на этой стадии играет фасилитатор либо формальный лидер, который должен решать внутригрупповые конфликты.

Стадия «бурления» приводит коллектив в состояние, когда его члены должны приложить усилия, чтобы совместить разнообразие идей со своими собственными убеждениями и ценностями. В этой зоне члены коллектива приспособляются к различиям во взглядах и сотрудничают друг с другом, разрабатывают групповые нормы поведения, возникает чувство товарищества и сплоченности.

Доработка предложений и подготовка альтернатив – члены коллектива переформулируют ценные мысли в конкретные предложения, и «шлифуют» их пока все участники дискуссии не придут к конечному решению, воплощающему все разнообразие точек зрения членов коллектива. Эта стадия характеризуется «сходящимся мышлением»: классификацией идей, их обобщением, вынесением оценок, в противовес расходящемуся, свойственному для первых стадий, в рамках которых поощряется открытая безоценочная дискуссия и генерация большого числа решений [3].

Принятие решения и расформирование. Выполняется принятие коллективного решения, учитывающего мнения всех экспертов, принявших участие в обсуждении, если коллектив был собран для решения единственной задачи, он расформировывается. Постоянные коллективы достигают этой стадии после реализации всех их целей.

Эти стадии согласуются с моделью «ромба группового принятия решений» (рис. 1), предложенной С. Кейнером, К. Толди, С. Фиск, Д. Бергером [3] для обнаружения и использования скрытых аспектов повседневной жизни коллектива, для диагностики, дорожной карты или обучающего инструмента, дающего коллективу единый язык

дискуссии. Модель показывает, что непонимание друг друга – часть коллективного принятия решения, а «зона бурления» – следствие разнообразия мнений в коллективе. Благодаря этому коллективам проще выдержать стресс «зоны бурления» и найти общую почву для совместной работы и принятия устраивающего всех решения.

Используя эту модель фасилитатор, отвечающий за успешность и эффективность коммуникации экспертов в коллективе, идентифицирует сложившуюся ситуацию принятия решений и пытается направить дискуссию в требуемое русло, активируя соответствующий стиль мышления в группе, например, предоставляя отдельным участникам приоритет в высказывании своих предложений. Рассмотрим подробнее первый этап модели «ромба группового принятия решений» – зону дивергентного мышления.

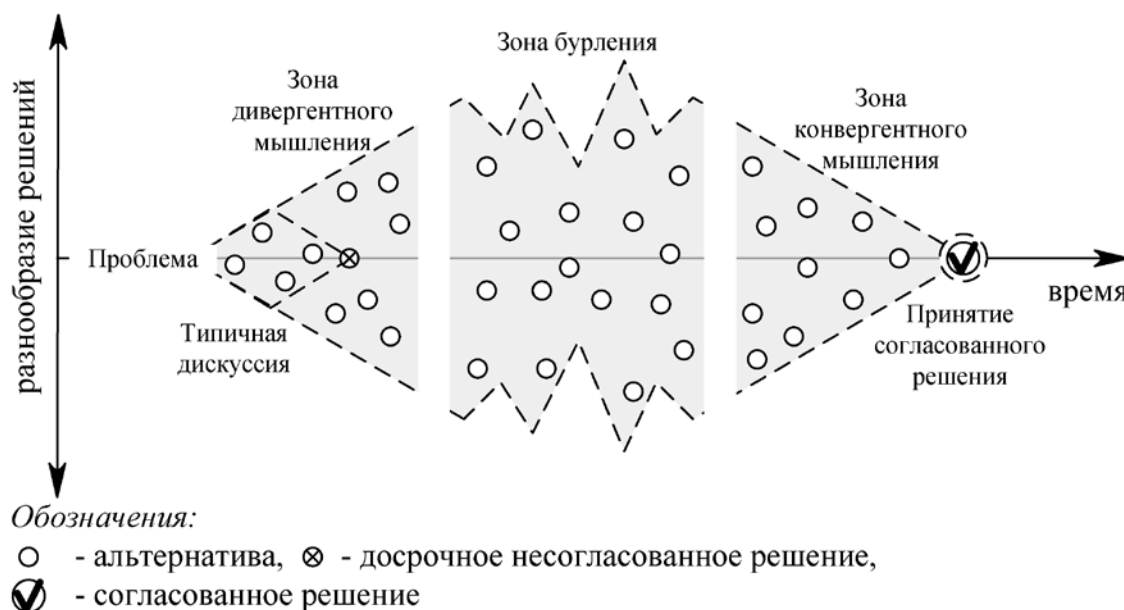


Рис. 1. Графическое представление модели ромба группового принятия решений С. Кейнера, К. Толди, С. Фиск, Д. Бергера.

Дивергентное мышление [10 – 12] – творческое мышление, «веерообразное», расходящееся, ассоциативное решение проблем. Оно характеризуется тем, что психологи называют беглостью восприятия, способностью генерировать несколько идей, гибкостью, способностью переходить на другую точку зрения и оригинальностью, способностью вырабатывать нетривиальные идеи.

Для запуска дивергентного мышления отдельного эксперта могут создаваться списки наводящих вопросов, выделяться время для размышлений и медитации, применяться метод ментальных карт, ведение журнала, создание иллюстраций и свободное письмо или рисование, когда человек сосредотачивается на конкретной теме и пишет или создает эскизы без остановок в течение короткого периода времени в потоке сознания [13].

Выделяются следующие наиболее известные подходы к организации дивергентного мышления в малых коллективах экспертов, которые, по сути, – формы мозгового штурма:

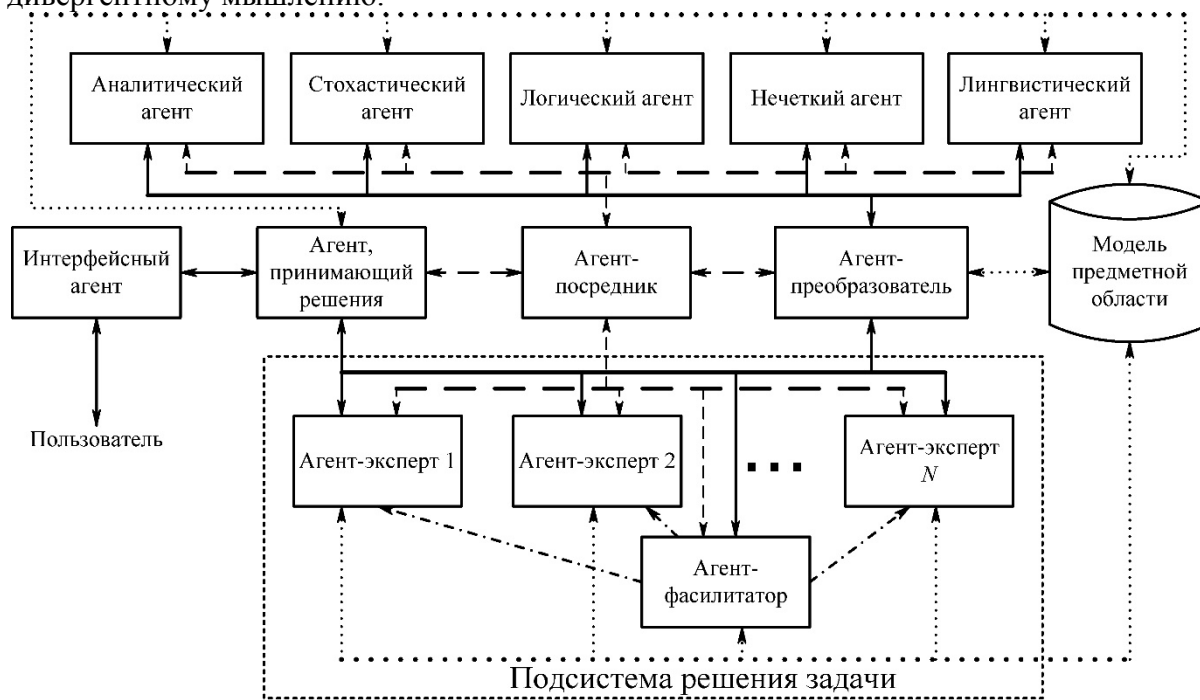
- Классический мозговой штурм – создание списка идей в творческом,

неструктурированном виде с целью получения большого числа идей за короткий промежуток времени. Ключевой инструмент – комбинирование или использование одной идеи для порождения новых. Во время мозгового штурма все идеи записываются, а не критикуется [14].

– Метод «6 идей в 1 идею» – разновидность мозгового штурма, комбинирующая дивергентное и конвергентное мышление. За короткое время, например, 10 мин каждый участник записывает пришедшие на ум 6 подходов к решению поставленной задачи, после чего предложения каждого участника обсуждаются. Каждый эксперт, получив мнения других участников о своих вариантах решения проблемы, дорабатывает одну из своих альтернатив. Доработанные варианты решения проблемы каждого участника формируют пул идей, с которыми группа работает на следующих этапах процесса принятия решения (рис. 1).

– Метод ментальных карт состоит в отображении предложений или идей в виде визуальной карты или картины, которая показывает отношения между ними. Создание ментальной карты начинается с отображения решаемой проблемы, затем показываются ответвления, представляющие собой разные задачи, части или аспекты основной проблемы. В результате создается визуальный образ или «карту» проблемы, которую коллектив может использовать для дальнейшего обсуждения проблемы [13].

– Мозговой штурм с наводящими вопросами – экспертам, кроме постановки возникшей проблемы, предоставляется список наводящих вопросов, например, «Как сейчас решается проблема?», «Кем она решается?», «Где и когда она решается?», «Что для этого нужно?», «Почему так?», «Как, кем, где и когда еще она может решаться?». Эти вопросы помогают экспертам лучше понять возникшую проблему, а последний вопрос заставляет рассмотреть различные варианты ее решения и побуждает к дивергентному мышлению.



Обозначения:

—> - отношения агентов: запросы информации, передача результатов их решения,

- - -> - отношения агентов: запросы помощи в решении задач,

.....> - взаимодействие (получение сведений из модели, обновление модели) агентов с моделью предметной области,

- - - -> - отношения фасилитации

– Рис. 2. Функциональная структура гибридной интеллектуальной многоагентной системы гетерогенного мышления.

Для компьютерного моделирования методов дивергентного мышления в малом коллективе экспертов предлагается функциональная структура гибридной интеллектуальной многоагентной системы гетерогенного мышления (ГИМСГМ), представленная на рис. 2.

Данная структура – гибридная интеллектуальная многоагентная система [15], расширенная агентом-фасилитатором, организующим коллективные процессы, и соответствующими отношениями между ним и агентами подсистемы решения задачи. Интерфейсный агент отвечает за взаимодействие с пользователем: запрашивает входные данные и выдает результат, а также визуализирует процесс решения проблемы. Агент, принимающий решения, ставит задачи агентам-экспертам, собирает результаты работы, определяет, достигнут ли критерий останова, и либо принимает итоговое решение, либо запускает новую итерацию процесса решения проблемы. Агенты-эксперты, имея знания о соответствующей своей «специальности» предметной области и моделируя определенный вид мышления, выполняют генерацию решений. Агент-посредник выполняет служебную работу по отслеживанию имен, моделей и возможностей зарегистрированных агентов. Агенты-решатели в верхней части рис. 2 вместе с агентом-преобразователем реализуют гибридную составляющую ГИМСГМ, комбинируя разнородные знания, методы, модели и алгоритмы решения задач. Модель предметной области – семантическая сеть, основа взаимодействия агентов, строится по концептуальной модели решаемой задачи.

Ключевое отличие ГИМСГМ от гибридных интеллектуальных многоагентных систем [15] – наличие агента-фасилитатора, который отвечает за организацию эффективной коллективной работы агентов: идентифицирует этапы процесса решения проблемы, состав агентов-экспертов и их «стиль мышления», сложившуюся ситуацию в ГИМСГМ, возникающие положительные и отрицательные групповые эффекты, воздействует на агентов-экспертов, чтобы активировать релевантные ситуации «стиль мышления», минимизировать отрицательные эффекты и усилить положительные, используя, среди прочего, модель «ромб группового принятия решений» (рис. 1). Для управления коллективной работой агентов ГИМСГМ агент-фасилитатор должен иметь нечеткую базу знаний о релевантности «стилей мышления» агентов различным ситуациям принятия решений в ГИМСГМ, а также перечисленных выше подходов дивергентного мышления возникающим проблемам. Для формирования такой базы знаний необходимо провести серию вычислительных экспериментов по решению проблем из различных классов [1] и установить соответствия между классом проблем и релевантным подходом к организации дивергентного мышления.

Предлагаемый класс ГИМСГМ обеспечивает комбинирование репрезентации неоднородной функциональной структуры сложной задачи (проблемы), с гетерогенным коллективным мышлением интеллектуальных агентов, что создает условия решать проблемы без упрощения и идеализации в условиях динамической среды.

ГИМСГМ по аналогии с многоагентными системами обладают высокой гибкостью, могут быть дополнены и модифицированы, устойчивы к сбоям, благодаря достаточному запасу компонентов и самоорганизации, имеют развитые возможности и средства адаптации к изменениям среды, в том числе путем модификации своей структуры и параметров. Применение ГИМСГМ в виртуальных коллективах, когда искусственные интеллектуальные агенты работают совместно с реальными экспертами, позволит поддерживать совместное принятие решений в смысле С. Кейнера и повышать групповую эффективность: качество решений; ответственность за принимаемые решения; значительно сокращать время реализации решений; улучшать отношения в группе; усиливать личную удовлетворенность участников; способствовать организационному обучению.

Рассмотрен процесс решения проблем малым коллективом экспертов и его основные стадии, выделены релевантные им стили мышления участников. Представлены основные подходы к организации первой стадии процесса решения проблем малым коллективом экспертов – зоны дивергентного мышления. Для релевантного моделирования процесса решения проблем малым коллективом экспертов и, в частности, процессов дивергентного мышления предложена функциональная модель ГИМСГМ. Применение предлагаемой модели позволяет соединять в информационной системе объективные (по крайней мере, относительно объективные) описания процессов и явлений во внешнем мире и субъективные подходы к восприятию во внутреннем мире коллектива интеллектуальных агентов. Субъективные знания аппроксимируют нелинейности, неоднородности, которыми изобилует внешний мир, поддерживают совместное в группе дивергентное мышление для выработки поля альтернатив, а объективные знания поддерживают конвергентное мышление над этим полем, критику, оценки и суждения для преодоления ограниченности теории принятия решений в динамичной среде.

Список литературы

1. Колесников А.В. Гибридные интеллектуальные системы. Теория и технология разработки / А.В. Колесников. — СПб.: Изд-во СПбГТУ, 2001. — 711 с.
2. Тарасов В.Б. От многоагентных систем к интеллектуальным организациям: философия, психология, информатика / В.Б. Тарасов. — М.: Эдиториал УРСС, 2002. — 352 с.
3. Kaner S. The Facilitator's Guide to Participatory Decision-Making / S. Kaner, L. Lind, C. Toldi, et. al. — San Francisco, CA: Jossey-Bass, 2011. — 368 p.
4. Боно Э. Параллельное мышление. От сократовского мышления к дебоновскому / Э. Боно. — Минск: Попурри, 2007. — 320 с.
5. Gardner H. Multiple Intelligences. The Theory in Practice. / H. Gardner. — New York: Basic Books, 1993. — 320 p.
6. Спиридонов В.Ф. Психология мышления. Решение задач и проблем / В.Ф. Спиридонов. — М.: Генезис, 2006. — 319 с.
7. Самсонова М.В. Технология и методы коллективного решения проблем: Учебное пособие / М.В. Самсонова, В.В. Ефимов. — Ульяновск: УлГТУ, 2003. — 152 с.
8. Занковский А.Н. Организационная психология: Учебное пособие для вузов по специальности «Организационная психология» / А.Н. Занковский. — М.: Флинта: МПСИ, 2002. — 648 с.
9. Организационное поведение: Учебник для вузов / Под ред. Г.Р. Латфуллина, О.Н. Громовой. — СПб: ЗАО Издательский дом «Питер», 2004. — 432 с.
10. Torrance E.P. Developing creative thinking through school experiences / E.P. Torrance // A source book for creative thinking. — New York: Charles Scribner's Sons, 1962. — P. 31 – 47.
11. Гилфорд Дж. Три стороны интеллекта / Дж. Гилфорд // Психология мышления. — М.: Прогресс, 1965. — С. 433 – 456.
12. Taylor C.W. Various approaches to and definitions of creativity / C.W. Taylor // The nature of creativity. — Cambridge: Cambridge Press, 1988. — P. 99 – 126.
13. Anoiko. Creativity. E-print: https://oiko.files.wordpress.com/2011/03/2011_wiki_anoiko_creativity1.pdf.
14. Strategies of Divergent Thinking. E-print: <http://faculty.washington.edu/ezent/imdt.htm>.
15. Колесников А.В. Решение сложных задач коммивояжера методами функциональных гибридных интеллектуальных систем / А.В. Колесников, И.А. Кириков, С.В. Листопад и др. — М.: ИПИ РАН, 2011. — 295 с.

УНИВЕРСАЛЬНЫЙ АППРОКСИМАТОР НА ОСНОВЕ МГУА-МОДЕЛЕЙ ДЛЯ ФОРМИРОВАНИЯ ВИРТУАЛЬНЫХ ПОТОКОВ

Т.В. Петрова, С.А. Филист, В.В. Жилин

Юго-Западный государственный университет, кафедра биомедицинской
инженерии, *SFilst@gmail.com*

При построении решающих модулей по принятию решений целесообразно вводить дополнительные информативные признаки, которые несут информацию о скрытых связях между исходными признаками [1, 2]. Вектор, соответствующий этим информативным признакам, назовем виртуальным потоком. Для двухальтернативных выборок необходимо построить две модели виртуальных потоков. Скрытые связи определяются аппроксимирующей функцией, построенной по данным, извлеченным из обучающей выборки (независимым переменным) и зависимым переменным, полученным посредством вероятностного программирования [2]. Если такая аппроксимирующая функция будет построена, то по данным неизвестного образца может быть найден дополнительный признак как функция этих входных данных.

Полагаем, что неизвестный образец характеризуется вектором наблюдаемых информативных признаков X с компонентами $x_1, x_2, \dots, x_m$. Необходимо получить функциональную зависимость латентной переменной Y от наблюдаемого вектора X , то есть

$$Y = F(x_1, x_2, \dots, x_m), \quad (1)$$

полагая, что присутствие на входе классифицирующей подели этого информативного признака улучшит качество классификации.

Модель нейросетевой структуры с виртуальным потоком показана на рис. 1 [3].

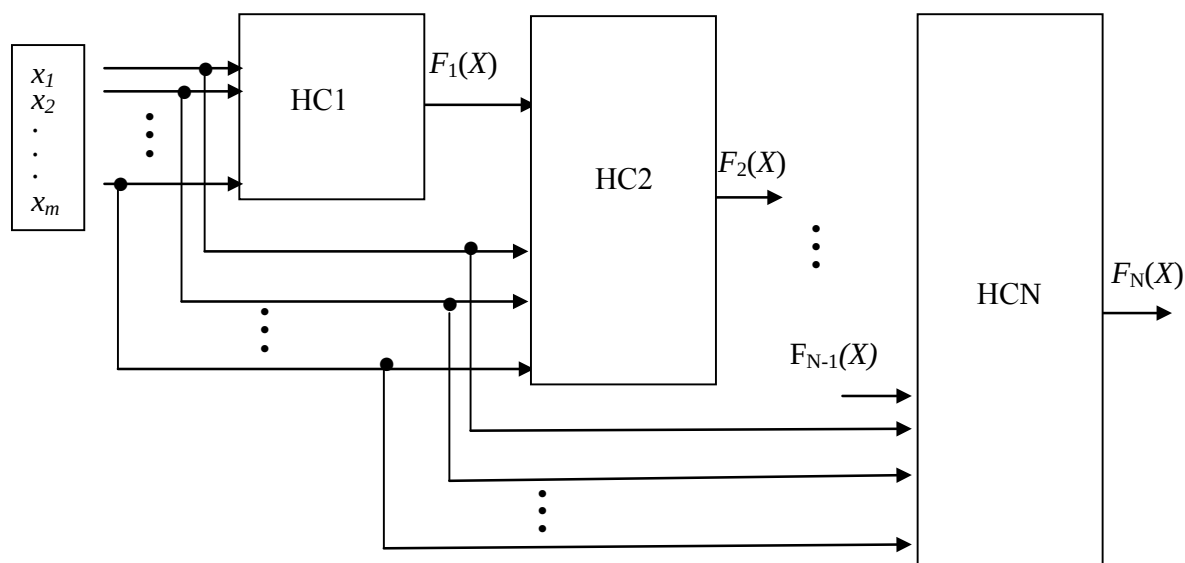


Рис. 1 - Структура нейросетевой модели с виртуальными потоками

При этом в обучающую выборку для аппроксиматоров входят не только экспериментальные данные, полученные на соответствующем интервале наблюдения,

но и данные, полученные в результате имитационного моделирования на предшествующей нейросетевой модели. То есть, каждый нейросетевой модуль (НС на рис. 1) в этой модели, кроме первого и последнего, является не только анализатором латентного параметра X , но источником данных для настройки параметров (обучения) последующего модуля.

Таким образом, необходимо решить задачу, состоящую в обнаружении и моделировании некоторой закономерности (1). Проблемы, возникающие при восстановлении функциональных зависимостей в условиях коротких выборок, являются типичными проблемами, возникающими при применении индуктивных методов, и существенно отличаются от классических проблем восстановления по выборкам большого объема. Особенность состоит в том, что при ограничении объема выборки качество восстановления зависит не только от качества аппроксимации в точках u_j , но еще и от таких факторов, как сложность аппроксимирующей функции и размерности пространства n .

Эта особенность заставляет сосредоточить внимание на правильном соотношении сложности приближающей функции с объемом обучающей выборки, так как имеющейся информации может не хватить даже для восстановления функции только в точках u_j . Тем более этой информации может не хватить для удовлетворительного восстановления в любой точке ее существования.

Среди немногих методов, в которых особое внимание уделяется поиску такого соотношения, выделяются метод группового учета аргументов (МГУА) [4].

Рассмотрим процесс синтеза модели оптимальной сложности более подробно.

В качестве такого полинома (1) очень часто используется полином Колмогорова-Габора [5]:

$$y = \alpha_0 + \sum_{i=1}^m \alpha_i x_i + \sum_{i=1}^m \sum_{j=1}^m \alpha_{ij} x_i x_j + \sum_{i=1}^m \sum_{j=1}^m \sum_{k=1}^m \alpha_{ijk} x_i x_j x_k + \dots, \quad (2)$$

так как с помощью такого полинома можно добиться достаточно точной аппроксимации любой дифференцируемой функции F .

Эта сложная зависимость заменяется множеством простых функций:

$$y_1 = f(x_1, x_2); \quad y_2 = f(x_1, x_3), \dots, y_s = f(x_{m-1}, x_m), \quad (3)$$

где $s = C_m^2$, причем функция f повсюду одинакова.

Очень часто в качестве функции f выбираются простые зависимости

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j \quad (4)$$

или

$$y(x_i, x_j) = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2, \quad (5)$$

связывающие только две переменные. Коэффициент этих зависимостей можно определить по МНК, имея соответственно 4 или 6 точек наблюдений в обучающей последовательности. Среди моделей первого ряда выбираются несколько, например S^* наилучших, показавших хорошие результаты на проверочной выборке. Среди отобранных моделей остаются только те, которые «впитали» в себя нечто большее, чем хорошая аппроксимация в узлах интерполяции; они «угадывают» поведение функции (2) в области, не охваченной экспериментом. Во втором ряду алгоритма полученные на обучающей выборке значения y_i , соответствующие отобранным моделям, рассматриваются в качестве аргументов нового ряда:

$$z_1 = f(y_1, y_2); \quad z_2 = f(y_1, y_3); \dots; \quad z_{S^*} = f(y_{S^*-1}, y_{S^*}). \quad (6)$$

Здесь функция f остаются прежними и соответствуют соотношениям (4) или (5), но аргументами этих функций выступают переменные y_i . Коэффициенты новых моделей (4.28) находятся по МНК на точках той же обучающей последовательности. Новые модели проверяются на точках проверочной последовательности, и среди них выбирается S^* наилучших, которые используются в качестве аргументов следующего третьего ряда и т.д. Сложность полиномов возрастает от ряда к ряду.

Входные аргументы и промежуточные переменные сопрягаются попарно, и сложность комбинаций на каждом ряду обработки информации возрастает (как при массовой селекции), пока не будет получена единственная модель оптимальной сложности.

Каждое частное описание является функцией только двух аргументов. Поэтому его коэффициенты легко определить по данным обучающей последовательности при малом числе узлов интерполяции. Исключая промежуточные переменные (если это удастся), можно получить аналог полного описания. Математика не запрещает обе эти операции. Например, по десяти узлам интерполяции можно получить в результате оценки коэффициентов полинома сотой степени и т.д.

Из ряда в ряд селекции пропускается только некоторое количество самых регулярных переменных. Степень регулярности оценивается по величине среднеквадратичной ошибки (средней для всех выбираемых в каждом поколении переменных или для одной самой точной переменной) на отдельной проверочной последовательности данных. Иногда в качестве показателя регулярности используется коэффициент корреляции.

Ряды селекции наращиваются до тех пор, пока регулярность повышается. Как только достигнут минимум ошибки, селекцию, во избежание «инцухта», следует остановить. Практически рекомендуется остановить селекцию даже несколько раньше достижения полного минимума, как только ошибка начинает падать слишком медленно. Это приводит к более простым и более достоверным уравнениям [5].

Среди множества МГУА - моделей выбираем L наилучших, которые могут быть представлены в виде множества

$$\{f_\ell(X_k)=\tilde{y}\}, \quad (7)$$

где $X_k \subseteq X$, $X=(x_1, x_2, \dots, x_m)$ – множество информативных признаков используемых решающими модулями; $\ell = \overline{1, L}$.

На рис. 2 представлена обобщенная структурная схема классифицирующего модуля, предназначенного для работы с дополненным на основе МГУА – моделей пространством информативных признаков.

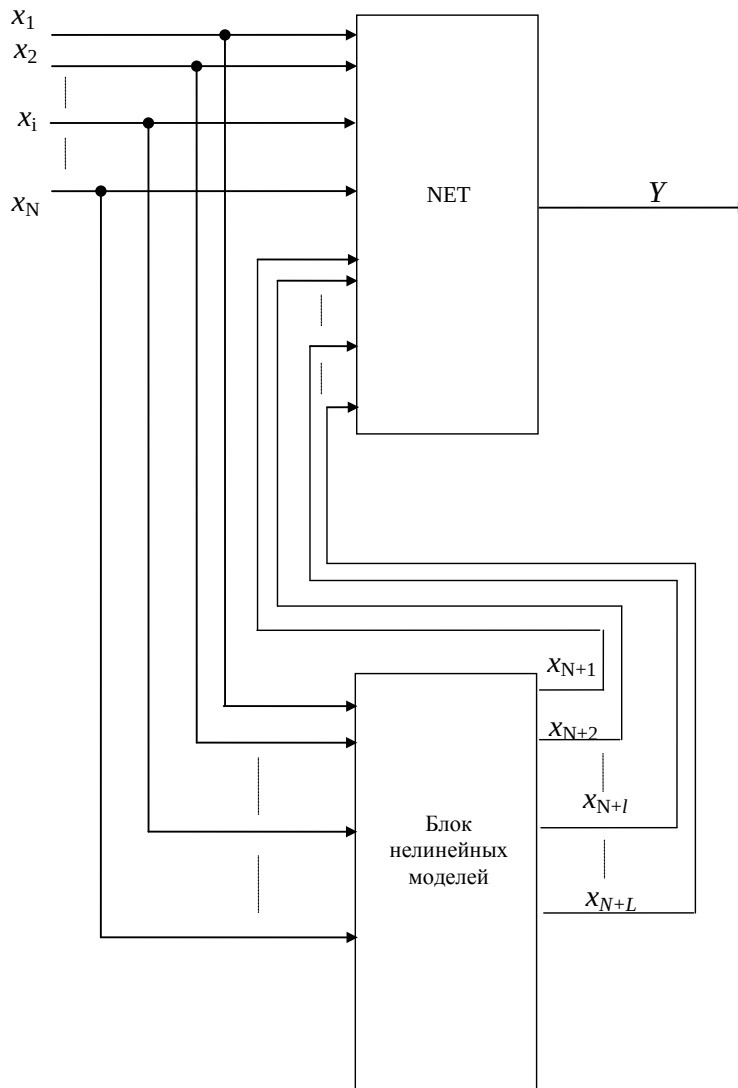


Рис. 2 - обобщенная структурная схема классифицирующего модуля, предназначенного для работы с виртуальными потоками, полученными на основе МГУА – моделей исходного пространства информативных признаков

Рассмотрим метод формирования множества (7). Метод должен синтезировать множество аппроксимирующих связей между элементами входного вектора X и позволить выбрать из этого множества L функциональных связей, которые формируют L дополнительных признаков, включение которых во входной вектор приводит к повышению качества классификации решающего модуля.

В случае вырождения нелинейных моделей в линейные многомерные аппроксиматоры блок нелинейных моделей (рис. 2) включает множество пар (в случае двухальтернативной классификации) классифицирующих функций (аппроксиматоров), каждый из которых выдает число, соответствующее состоянию входного вектора. При необходимости, эти два числа (a и b ,) могут быть агрегированы в одно посредством одного из следующей группы агрегаторов:

$$x_{N+i} = \max(a, b); \quad x_{N+i} = a - b; \quad x_{N+i} = \min(a, b); \quad (8)$$

$$x_{N+i} = \begin{cases} 0, & \text{при } a \leq b; \\ 1, & \text{при } a > b. \end{cases}$$

При этом необходимо учитывать, что соответствующие выходы блоков, моделирующие корреляционные связи исходных информативных признаков, могут быть как двухкомпонентными векторами, так и скалярами, в зависимости от наличия в моделирующих модулях агрегирующих блоков вида (8).

Включение в множество моделей очередной, $(\ell + 1)$ -й, модели, осуществляется по рекуррентной схеме, представленной на рис. 1. Эта схема позволяет оценить вклад в показатели качества принятия решений $(\ell + 1)$ -го дополнительного информативного признака при наличии (ℓ) дополнительных признаков.

В общем случае, для построения модели рис. 2 привлекаются алгоритмы эволюционного типа («жадные» алгоритмы) [5].

Предложен метод формирования нелинейных моделей виртуальных потоков, составляющих основу классифицирующего модуля рис. 2. Модуль моделирования виртуальных потоков состоит из двух слоев. Первый слой формирует множество моделей $\hat{Y}$. Для каждого виртуального потока посредством МГУА - моделирования получено свое подмножество моделей $\{z_j\}_i \subseteq Z, \forall i = \overline{1, M}$. Каждое подмножество моделей $\{z_j\}_i, j = \overline{1, K_i}$, где K_i – число МГУ - моделей для i -го виртуального потока, полученных на основе МГУА-моделирования, которые предполагается использовать для описания взаимного влияния известных информативных признаков (реальных потоков) в системе простых комбинаций реальных и виртуальных потоков. Множество статических моделей $\hat{Y}$ получается не посредством усложнения МГУА-моделей, а посредством МГУА-нейронной сети [4]. Отличие предлагаемой МГУА-нейронной сети от рассмотренной в [4] состоит в том, что каждый блок модели виртуального потока (второй слой модели) является МГУА-нейронной сетью, на входы которой поступают полученные путем МГУА-моделирования модели, включающие множество X реальных и виртуальных потоков. Структурная схема i -го блока модели виртуального потока представлена на рис. 3.

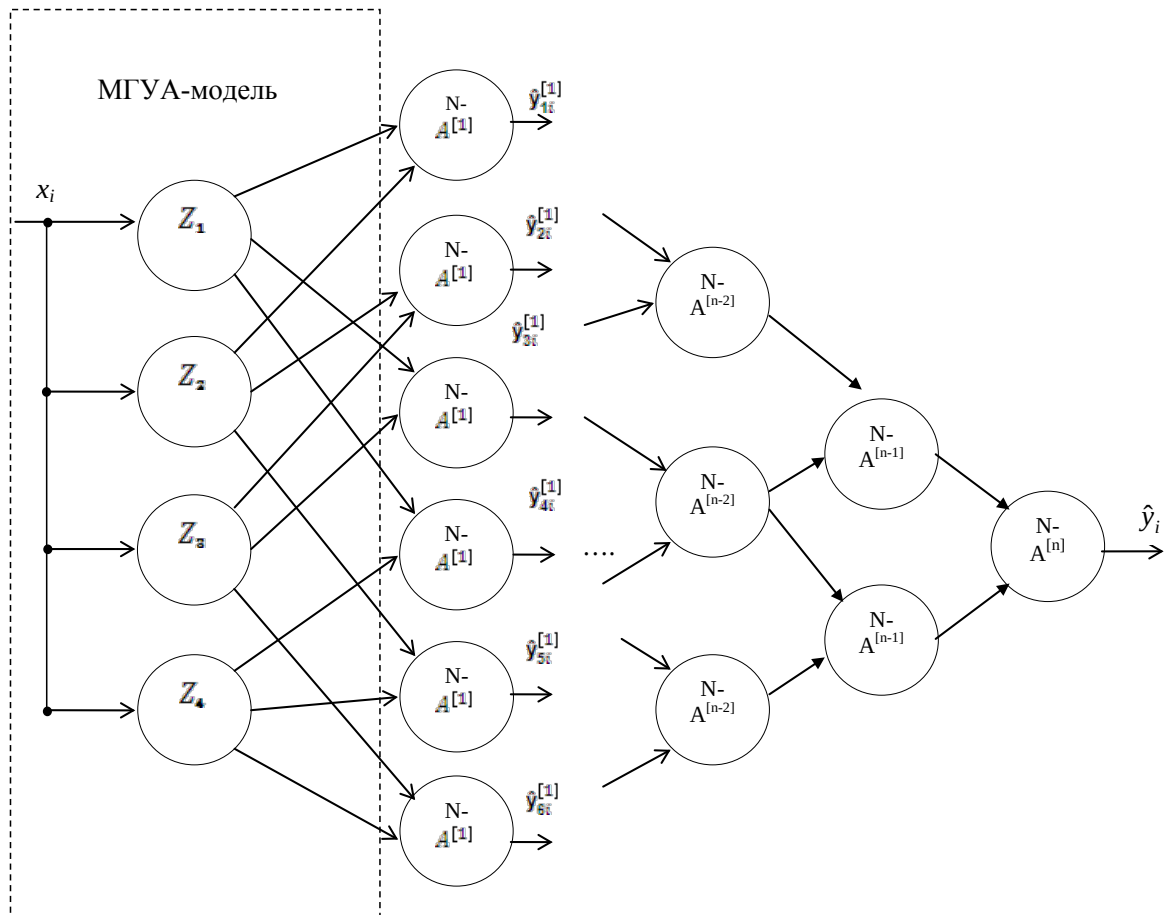


Рис. 3 – Структурная схема i - й нелинейной статической модели первого слоя нелинейной динамической модели на основе МГУА – моделирования

В качестве примера в ней использованы четыре МГУА - модели виртуального потока $x_i: z_1, z_2, z_3, z_4$ с нелинейными адалинами (N-A).

МГУА - сеть имеет переменную структуру, которая может изменяться в процессе обучения. Каждый нейрон сети – N-адалина представляет собой адаптивный линейный ассоциатор с двумя входами z_g и z_h и нелинейным процессором, образованным тремя блоками умножения, и вычисляет квадратичную комбинацию входов вида

$$\begin{aligned} \hat{y}_{gh} = f_{gh}(z_g, z_h) = & w_{gh_0} + w_{gh_1} z_g + w_{gh_2} z_g^2 + w_{gh_3} z_g z_h + \\ & + w_{gh_4} z_h^2 + w_{gh_5} z_h = w_{gh}^T z_{gh}, \end{aligned} \quad (9)$$

где $w_{gh} = (w_{gh_0}, w_{gh_1}, w_{gh_2}, w_{gh_3}, w_{gh_4}, w_{gh_5})^T$, $z_{gh} = (1, z_g, z_g^2, z_g z_h, z_h^2, z_h)^T$.

Процесс обучения МГУА - модели виртуального потока состоит в конфигурации нейронных сетей (рис. 3), начиная с первого скрытого слоя, независимой настройке синаптических весов каждой нелинейной адалины и наращивании количества слоев для достижения необходимой точности прогнозирования. Количество нейронов первого скрытого слоя сети определяется количеством K МГУА – моделей соответствующего

эндогенного фактора и не превышает значение $K(K-1)/2$ – количества сочетаний из K по 2.

После обучения нейронной сети посредством любого из известных алгоритмов обучение [6] оценивается точность моделирования, например, с помощью дисперсии ошибки предсказания для каждого нейрона и формируется группа нейронов, дающих ошибку ниже некоторого априорно заданного порога. Именно выходы этой группы $\hat{y}_i^{[1]}$ являются входами нейросетевой структуры NET1 [4].

Список литературы

1. Позин, А.О. Синтез дополнительного информативного признака для нейронной сети прямого распространения / А.О. Позин, С.А. Филист, А.Н. Шуткин и др. // Новые решения в области упрочняющих технологий: взгляд молодых специалистов: сборник научных статей по материалам Международной научно-практической конференции. – Курск: ЗАО «Университетская книга», 2016. – С. 212 - 217.

2. Позин, А.О. Тестирующие гибридные системы с дополнительным пространством информативных признаков / А.О. Позин, С.А. Филист, А.Н. Шуткин // Современные информационные технологии в управлении качеством: сборник статей V Международной научно-прикладной конференции. – Пенза: Приволжский Дом знаний, 2016. – С.46-50.

3. Старцев, Е.А. Двумерные нейросетевые структуры мета-анализа медико-экологических данных /Е.А. Старцев, О.В. Шаталова, В.В. Уварова//Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века: сб. ст. по материалам Всерос. науч.-практ. конф. – Пермь: Перм. гос. нац. исслед. ун-т., 2016. – С.161-164.

4. Позин, А.О. МГУА-нейронные сети для прогнозирования состояния сложных систем с временными лагами /А.О. Позин, Е.А. Старцев, В.В. Уварова//Нейроинформатика, ее приложения и анализ данных: Материалы XXIV Всероссийского семинара. – Красноярск: Институт вычислительного моделирования СО РАН. – 2016. – С. 50-55.

5. Артеменко, М.В. Формирование множества информативных показателей на основании аппроксимирующего полинома Колмогорова-Габора и максимального градиента функциональных различий /М.В. Артеменко, Н.М. Калугин, А.Н. Шуткин //Известия Юго-Западного государственного университета. Серия Управление, вычислительная техника, информатика. Медицинское приборостроение. - 2016, №1(18), с. 116-123.

6. Хайкин С. Нейронные сети: полный курс. 2-е изд., испр.: Пер. с англ. – М.: ООО «И.Д. Вильямс», 2006.- 1104 с.

ГЛОБАЛЬНАЯ ОПТИМИЗАЦИЯ НА ОСНОВЕ НЕЙРОСЕТЕВОЙ АППРОКСИМАЦИИ ИНВЕРСНЫХ ЗАВИСИМОСТЕЙ С ЭВОЛЮЦИОННОЙ НАСТРОЙКОЙ ПАРАМЕТРОВ

К.В. Пушкарев

Сибирский федеральный университет,
Институт космических и информационных технологий,
каф. вычислительной техники, *kpushkaryov@gmail.com*

Введение

Глобальная оптимизация целевых функций (ЦФ) многих переменных, рассматриваемых как «чёрные ящики» при минимальных требованиях к свойствам функции, является актуальной проблемой, поскольку обещает универсальный подход к решению многих практических задач (инженерных, машинного обучения и т. д.).

Хотя при абсолютном отсутствии требований к ЦФ задача практически неразрешима [1], представляет интерес поиск таких работающих в более узком классе задач методов, которые извлекали бы максимальную пользу из той информации, которую они получают о ЦФ, вычисляя её значения в пробных точках. Важным шагом к этой цели видится долговременное сохранение имеющей значение для поиска глобального минимума информации о ЦФ в виде параметров, влияющих на поведение метода оптимизации. Естественно желать, чтобы адаптация метода к задаче происходила посредством самонастройки.

Метод нейросетевой аппроксимации инверсных зависимостей

Эвристический метод нейросетевой аппроксимации инверсных зависимостей (НАИЗ) [2] предназначен для нахождения глобального минимума непрерывных целевых функций многих переменных $\Phi(\mathbf{x})$ в области, имеющей вид многомерного параллелепипеда ($\Omega = \{\mathbf{x} : L_i \leq x_i \leq U_i, i = \overline{1, n}\}$). ЦФ рассматривается как «чёрный ящик».

В основе метода лежит итеративное понижение значения целевой функции с отображением его в координаты с помощью обобщённо-регрессионных нейронных сетей (Generalized Regression Neural Network, GRNN), аппроксимирующих инверсные зависимости (координат от значения ЦФ). Базовый алгоритм состоит в следующем:

1. Инициализация: $k := 1$; $\Phi(\mathbf{x})$ — начальное множество пробных точек, представляющих собой точки из области поиска с вычисленными в них значениями ЦФ.

2. GRNN обучается отображать значения ЦФ в соответствующие координаты по множеству $P^{[k]}$.

3. Из множества $P^{[k]}$ выбирается точка с наименьшим значением ЦФ $\langle \mathbf{x}_{min}^{[k]}, \Phi(\mathbf{x}_{min}^{[k]}) \rangle$, это значение понижается на некоторую величину $\varepsilon_f^{[k]}$ (декремент) и отображается в координаты с помощью GRNN: $\mathbf{x}_{min}^{*[k]} = GRNN^{[k]}(\Phi(\mathbf{x}_{min}^{[k]}) - \varepsilon_f^{[k]}, spread^{[k]})$. В полученной точке вычисляется истинное значение ЦФ, в результате чего получается новая пробная точка $\langle \mathbf{x}_{min}^{*[k]}, \Phi(\mathbf{x}_{min}^{*[k]}) \rangle$.

4. Новая точка включается во множество $P^{[k]}$, если значение ЦФ в ней меньше текущего наименьшего:

$$P^{[k+1]} = \begin{cases} P^{[k]} \cup R^{[k]} \cup \left\{ \left\langle \mathbf{x}_{min}^{*[k]}, \Phi(\mathbf{x}_{min}^{*[k]}) \right\rangle \right\}, & \text{если } \Phi(\mathbf{x}_{min}^{*[k]}) < \Phi(\mathbf{x}_{min}^{[k]}); \\ P^{[k]} \cup R^{[k]}, & \text{иначе,} \end{cases}$$

где $R^{[k]}$ — множество пробных точек, созданных случайным образом.

5. $k := k + 1$.

6. Шаги 2-5 повторяются до достижения условий останова.

Параметр GRNN spread регулирует гладкость аппроксимации.

Применение вышеописанного метода на практике сопряжено с рядом трудностей. Его эффективность быстро падает с ростом числа переменных. В качестве стартовой на каждой итерации используется только одна (лучшая) пробная точка, а случайное вбрасывание является недостаточно эффективным источником дополнительных пробных точек, из-за чего поиск может зайти в тупик. Для преодоления этой проблемы метод нейросетевой аппроксимации инверсных зависимостей был включён в состав гибридного эвристического параллельного метода [3]. В гибридном эвристическом параллельном методе поиск глобального минимума выполняется оптимизационными агентами, которые итеративно запускают ряд алгоритмов оптимизации (поисковых инструментов), обменивающихся информацией через общую таблицу пробных точек.

Ещё одна проблема состоит в том, что когда ЦФ имеет несколько близких минимумов, то линия инверсной зависимости проходит посередине между ними. Для преодоления этого была предложена модификация GRNN — дуальная обобщённо-регрессионная нейронная сеть (DGRNN) [4]. У этой сети есть дополнительный вход, на который подаются координаты точки — фокуса поиска. Каждый пример при обучении GRNN/DGRNN состоит из входного и выходного эталона. Чем дальше от фокуса находится запомненный нейросетью эталон, тем ниже его активация. В результате эталоны, находящиеся вблизи фокуса, приобретают преимущество. DGRNN имеет дополнительный параметр argspread, являющийся аналогом spread для второго входа.

Эволюционный алгоритм настройки параметров метода

Чтобы ограничить диапазон изменения параметров, разности значений ЦФ между входом сети и эталонами, вычисленные внутри DGRNN, делятся на квантили абсолютных значений разностей порядка p_{f1} для отрицательных разностей и p_{f2} для положительных. Величины p_{f1} , p_{f2} становятся входными параметрами нейросети и заменяют собой spread. Расстояния по координатам делятся на собственный квантиль порядка p_x , который заменяет собой argspread. Декремент ЦФ ε_f применяется уже после этих операций, поэтому также выражается в относительных величинах и является третьим входным параметром.

Эволюционная настройка параметров основана на принципах генетического алгоритма. Параметры задаются правилами, которые скрещиваются, мутируют и проходят отбор.

Применение правила в базовой точке (БТ) $\mathbf{b}$ состоит в следующем. Сначала формируется присоединённое к БТ множество пробных точек:

$$P_b = \left\{ \langle \mathbf{x}, \Phi(\mathbf{x}) \rangle : \mathbf{x} = \mathbf{b} + i \cdot \mathbf{p}_j, i = -i_{\max}, i_{\max}, j = 0, j_{\max} \right\}$$

где $\mathbf{p}_j$ — j -й вектор шаблона данного правила.

По присоединённому множеству обучается модифицированная DGRNN, отображающая значения ЦФ в координаты. Значение ЦФ в БТ с помощью нейросети отображается в координаты:

$$\mathbf{b}^* = DGRNN(\Phi(\mathbf{b}), \mathbf{b}, \varepsilon_f, p_{f1}, p_{f2}, p_x).$$

При этом БТ используется в качестве фокуса поиска, а значения параметров ε_f , p_{f1} , p_{f2} , p_x задаются правилом. Далее значение ЦФ вычисляется в точке с координатами $\tilde{\mathbf{b}} = \mathbf{b} + \alpha_b(\mathbf{b}^* - \mathbf{b})$, где параметр α_b также задаётся правилом. Получившуюся точку $\langle \tilde{\mathbf{b}}, \Phi(\tilde{\mathbf{b}}) \rangle$ будем называть *кандидатом от правила*.

После применения правил для каждой БТ определяется *цель*. Цель БТ первоначально равна её начальному положению. Далее на каждой итерации в качестве нового значения цели БТ выбирается лучшая по значению ЦФ точка из следующего списка: предыдущее значение цели, лучший кандидат от правил на данной итерации, текущее положение БТ. Наконец, каждая БТ сдвигается в сторону цели по правилу, аналогичному используемому в методе роя частиц (PSO):

$$\begin{aligned} \mathbf{v}^{[k]} &= k_i \mathbf{v}^{[k-1]} + k_l \mathbf{U}_l \otimes (\mathbf{g}_l^{[k]} - \mathbf{b}^{[k]}) + k_g \mathbf{U}_g \otimes (\mathbf{g}_g^{[k]} - \mathbf{b}^{[k]}), \\ \mathbf{b}^{[k+1]} &= \mathbf{b}^{[k]} + \mathbf{v}^{[k]}, \end{aligned}$$

где k — номер итерации; k_i , k_l , k_g — числовые коэффициенты; $\mathbf{g}_l^{[k]}$ — собственная цель БТ; $\mathbf{g}_g^{[k]}$ — глобальная цель; $\mathbf{U}_l$, $\mathbf{U}_g$ — векторы равномерно распределённых случайных чисел, диапазон которых является параметром алгоритма; $\otimes$ — покомпонентное произведение векторов.

Глобальная цель $\mathbf{g}_g^{[k]}$ определяется как лучшая из собственных целей БТ группы (роя), к которой принадлежит данная БТ. К глобальной цели с установленной периодичностью применяется алгоритм локального поиска. В данной работе использовался модифицированный алгоритм BFGS [5, С. 136].

Таким образом, правило является носителем определённого набора значений параметров (генотипа): ε_f , p_{f1} , p_{f2} , p_x , α_b , $\{\mathbf{p}_j\}$. Приспособленность каждого правила выражается в виде двух числовых характеристик: *очарования (charm)* и *достоинства (merit)*.

К установленному количеству базовых точек с установленной периодичностью применяются все правила. Для каждой невыбранной на данной итерации БТ выбирается одно случайное правило так, что вероятность выбора пропорциональна rsc^i , где $0 < rsc < 1$ — параметр алгоритма, а i — ранг правила в ряду по убыванию очарования, отсчитывающийся от нуля.

В результате применения всех правил к каждой выбранной БТ каждому правилу сопоставляется вектор очков, i -й компонент которого — это количество раз, когда кандидат от данного правила занимал i -е место в общем списке кандидатов от всех правил для данной БТ, упорядоченном по убыванию прогресса кандидата (разности между значением ЦФ в БТ и у кандидата), причём все положительные прогрессы ставятся впереди отрицательных. Правила упорядочиваются по убыванию вектора очков в словарном порядке. После этого очарование правила устанавливается равным rrd^i , где $0 < rrd < 1$ — параметр алгоритма, а i — ранг правила в полученном ряду, отсчитывающийся от нуля.

Достоинство правила определяется по формуле:

$$merit^{[k+1]} = merit^{[k]} \cdot rmd^a + charm^{[k]} + ppe^{[k]},$$

где $rmd < 1$ — параметр алгоритма; a — возраст правила в итерациях; $ppe^{[k]}$ — отношение среднего прогресса данного правила на итерацию к максимальному среднему прогрессу на итерацию среди всех правил.

Вероятность выбора правила в качестве родителя для мутации и скрещивания пропорциональна его очарованию. Мутация заключается в случайных нормально распределённых возмущениях параметров правила. Скрещивание скалярных параметров происходит по формуле $\beta_o = r\beta_{p1} + (1-r)\beta_{p2}$, где β_o , β_{p1} , β_{p2} — значение параметра соответственно у потомства, первого и второго родителя, r — случайный равномерно распределённый коэффициент от 0 до 1, выбираемый для каждого параметра независимо. Скрещивание векторов шаблона $\{\mathbf{p}_j\}$ происходит путём случайного выбора пар векторов шаблонов первого и второго родителя, после чего соответствующие компоненты векторов скрещиваются как скаляры, как описано выше. Количество векторов шаблона потомства определяется биномиальным распределением $B(n, 0,5)$, где n — сумма размеров шаблонов родителей.

На каждой итерации создаётся установленное количество новых правил путём скрещивания и мутации старых, а также случайной генерации. Старые правила упорядочиваются по убыванию достоинства. Установленное количество первых правил из этого ряда сохраняется для перехода в следующую итерацию (элита). Популяция правил для следующей итерации формируется из элитных и новых правил.

Процесс продолжается до выполнения условий останова, таких как снижение среднего убывания лучшего значения ЦФ за итерацию ниже порога и др. После завершения процесса итоговая популяция правил сохраняется в виде базы правил, которая может быть загружена в начале нового процесса.

Эксперименты

Для проверки эффективности метода и подтверждения влияния адаптации были проведены эксперименты по решению тестовых задач минимизации функций 50 и 100 переменных. За основу были взяты классические тестовые ЦФ (табл. 1)

Таблица 1

Базовые целевые функции

Название	Выражение	Глобальный минимум
Функция Растригина	$f_1(\mathbf{x}) = \sum_{i=1}^n [x_i^2 - 10\cos(2\pi x_i) + 10]$	$f_1(\mathbf{0}) = 0$
Функция Розенброка	$f_2(\mathbf{x}) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$f_2(\mathbf{1}) = 0$
Функция Экли	$f_3(\mathbf{x}) = -20\exp\left(-0,2\sqrt{\frac{1}{n}\sum_{i=1}^n x_i^2}\right) - \exp\left(\frac{1}{n}\sum_{i=1}^n \cos(2\pi x_i)\right) + 20 + e$	$f_3(\mathbf{0}) = 0$

Ко всем функциям были применены одинаковые преобразования координат, заключающиеся в сдвиге и ортогональном преобразовании: $f_i^*(\mathbf{x}) = f_i(M(\mathbf{x} - \mathbf{x}_0))$. Компоненты вектора сдвига $\mathbf{x}_0$ случайны и равномерно распределены от -5 до 5 . Матрица M получена путём ортонормализации матрицы случайных чисел, равномерно распределённых от 0 до 1. Данные преобразования обусловлены тем, что для классических тестовых функций слабым местом является положение глобального минимума в нуле или другой точке с одинаковыми координатами, а также возможность искать минимум независимо по отдельным координатам. В результате показатели метода оптимизации, использующего эти особенности априори, не будут объективно отражать его применимость для решения задач общего вида той же размерности. Аналогичный

подход к созданию тестовых функций применён в [6].

Область поиска была установлена от -100 до 100 по каждой координате для функций f_1^* , f_3^* и от -50 до 50 для f_2^* . Уменьшение области для f_2^* объясняется тем, что рассматриваемый метод не находит глобальный минимум функции Экли в области от -100 до 100 . Это можно объяснить тем, что у функции Экли при большом удалении от глобального минимума начинает абсолютно преобладать колебательная составляющая, которая практически полностью скрывает глобальную тенденцию.

Было проведено два типа экспериментов. *Первый тип* был направлен на общую оценку эффективности метода. Количество переменных было равно 50. Для каждой ЦФ алгоритм запускался 100 раз со случайной инициализации. База правил также при каждом запуске инициализировалась случайно. *Второй тип* был направлен на выявление влияния долговременной адаптации правил. Для этого при 100 переменных было проведено 50 серий по 10 запусков алгоритма для каждой ЦФ со случайной инициализацией (кроме правил). База правил инициализировалась случайно в начале серии, дальше после каждого запуска она сохранялась и загружалась вновь в начале следующего запуска.

Результаты

Итоговая статистика найденных минимальных значений ЦФ (Φ_{min}) и количества вычислений ЦФ (N_{eval}) для экспериментов первого типа приведена в табл. 2.

Таблица 2

Результаты экспериментов первого типа

ЦФ	Параметр	Минимум	Максимум	Среднее	Медиана	Среднеквадратичное отклонение
1	Φ_{min}	0,0000	22,8841	0,8855	0,0000	3,9796
	N_{eval}	6 391 132	12 837 968	7 667 724	7 293 589	1 179 328
2	Φ_{min}	0,0000	3,9866	0,7973	0,0000	1,6027
	N_{eval}	2 969 366	5 206 920	4 365 121	4 595 910	648 184
3	Φ_{min}	0,0000	19,8916	2,0086	0,0000	5,9635
	N_{eval}	2 406 230	9 623 101	4 514 011	4 270 251	994 629

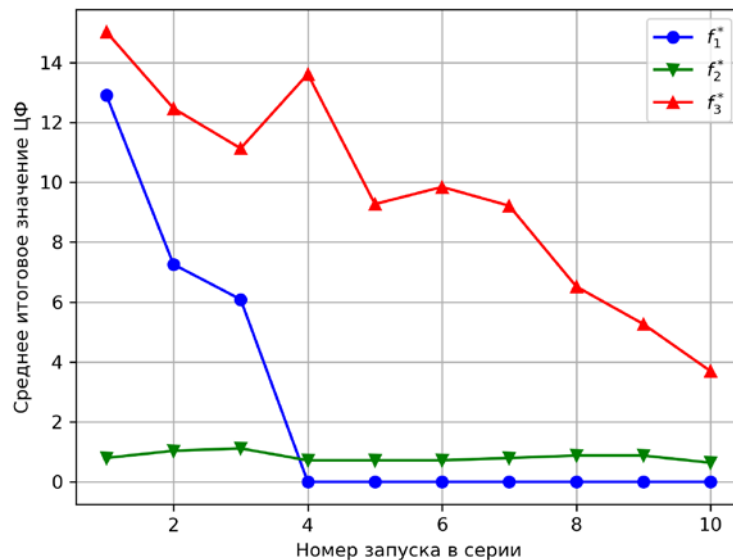
Результаты экспериментов второго типа показаны на рисунке. Сравнение выборок итоговых значений ЦФ для первого и десятого запусков для каждой ЦФ с помощью двустороннего знаково-рангового критерия Уилкоксона при уровне значимости 0,05 показало значимые отличия для f_1^* ($p = 0,001$), f_3^* ($p = 10^{-7}$) и отсутствие таковых для f_2^* ($p = 0,65$).

Выводы

По результатам экспериментов можно сделать вывод, что модифицированный метод нейросетевой аппроксимации инверсных зависимостей с эволюционной настройкой параметров способен находить глобальный минимум некоторых сложных многоэкстремальных целевых функций многих переменных.

Статистически значимое убывание в среднем итоговых значений ЦФ в режиме с

загрузкой последней базы правил говорит о постепенной адаптации метода к решаемой задаче. Отсутствие такой тенденции для функции f_2^* может объясняться тем, что для функции Розенброка локальный поиск, применяемый в рамках представленного метода, сам по себе очень эффективен, в результате чего эволюционная настройка параметров не играет существенной роли.



Среднее итоговое значение ЦФ в зависимости от номера запуска в серии в экспериментах второго типа.

Список литературы

1. Wolpert D. H. No free lunch theorems for optimization / D. H. Wolpert, W. G. Macready // IEEE transactions on evolutionary computation. — 1997. — Т. 1. — № 1. — С. 67-82.
2. Koshur V. Global optimization via neural network approximation of inverse coordinate mappings / V. Koshur, K. Pushkaryov // Optical Memory and Neural Networks. — 2011. — Vol. 20, no. 3. — Pp. 181–193. — URL: <http://dx.doi.org/10.3103/S1060992X1103009X>.
3. Пушкарев К. В. Гибридный эвристический параллельный метод глобальной оптимизации / К. В. Пушкарев, В. Д. Кошур // Вычислительные методы и программирование. — 2015. — Т. 16. — С. 242–255.
4. Кошур В Д. Дуальные обобщённо-регрессионные нейронные сети для решения задач глобальной оптимизации / В. Д. Кошур, К. В. Пушкарев // Научная сессия НИЯУ МИФИ — 2010. XII Всероссийская научно-техническая конференция «Нейроинформатика-2010»: Сборник научных трудов. В 2-х частях. Ч. 2 / Отв. ред. О. А. Мишулина. — М.: НИЯУ МИФИ, 2010. — С. 219-227.
5. Nocedal J. Numerical Optimization: Second Edition / J. Nocedal, S. J. Wright. — Springer-Verlag New York, 2006. — 664 p.
6. Liang J. Problem Definitions and Evaluation Criteria for the CEC 2013 Special Session on Real-Parameter Optimization: tech. rep. / J. J. Liang, B. Y Qu, P. N. Suganthan, A. G. Hernandez-Diaz. — 2013. — URL: http://www.ntu.edu.sg/home/EPNSugan/index_files/CEC2013/CEC2013.htm.

ДИСТАНЦИОННОЕ ОБУЧЕНИЕ ЛЮДЕЙ С ОСЛАБЛЕННЫМ ЗРЕНИЕМ НА ОСНОВЕ КОГНИТИВНОЙ SMART – ТЕХНОЛОГИИ³

Г.А. Самигулина¹, Ж.С. Лукманова²

¹ Институт информационных и вычислительных технологий КН МОН РК,
Казахстан, Алматы galinasamigulina@mail.ru

² Казахский национальный технический университет имени К. И. Сатпаева
azeshova@mail.ru

Четвертая промышленная революция кардинально изменила информационное пространство, в котором системы дистанционного обучения (ДО) поднимаются на новый уровень, основу которого составляют интеллектуальные подходы. Приоритетным направлением являются Smart – технологии, которые предполагают применение когнитивного подхода при организации процесса обучения [1]. В работе [2] описаны современные подходы искусственного интеллекта (ИИ) и когнитивные технологии для улучшения традиционных методов преподавания в онлайн режиме. Особый интерес [3] вызывают когнитивные Smart - технологии ДО для людей с ослабленным зрением (ЛОЗ). В статье [4] отмечается, что хорошее зрение является главным условием для успешного ДО. Проблемы со зрением приводят к снижению эффективности учебного процесса и требуют определенной коррекции поведения при изучении информации за компьютером [5]. Актуальна разработка специализированных систем ДО для людей с ослабленным зрением. Исследования [6] посвящены разработке мультиагентной системы дистанционного обучения для людей с ограниченными возможностями зрения на основе нейронных сетей и нейро-нечеткой логики. В статье [7] представлены исследования для разработки специальных подходов коммуникации ЛОЗ с помощью эффективных мультимедийных средств. В работе [8] рассмотрена технология, связанная с отслеживанием взгляда при прохождении тестов для определения психологического состояния человека.

В последнее время актуальна разработка интеллектуальных информационных технологий ДО с использованием различных нетрадиционных биоинспирированных подходов на основе нейронных сетей, генетических алгоритмов, искусственных иммунных систем (ИИС) [9-11], алгоритмов роевого интеллекта и т.д. Индивидуальные различия в характере обучающихся являются предметом современных исследований. В статье [2] рассматриваются различные психотипы человека, влияние их на характер и поведение. В исследованиях [13,14] на основе методов ИИ анализируется эмоциональная устойчивость обучающихся и её влияние на успех работы в онлайн-группах. При определении психотипа огромную роль играют эмоции человека, которые могут быть негативными или положительными. Каждый психотип человека имеет сильные и слабые стороны, но это не означает, что кто-то лучше или хуже. Каждый психотип по-разному воспринимает информацию и реагирует на окружающий мир [15-18].

³ Работа выполняется по гранту Комитета Науки Министерства Образования и Науки Республики Казахстан (2018–2020 гг.), по теме «Разработка когнитивной Smart – технологии для интеллектуальных систем управления сложными объектами на основе подходов искусственного интеллекта».

Предлагаемые исследования посвящены актуальной проблеме создания инновационной когнитивной Smart-технологии для дистанционного обучения (ДО) людей с ослабленным зрением инженерным специальностям. Применение когнитивного подхода позволяет обеспечить качественное персонализированное дистанционное обучение в зависимости от типа центральной нервной системы обучающихся, психофизиологических особенностей восприятия и усвоения текущей информации, а также особенностей зрения. Обработка дескрипторов, характеризующих каждого обучающегося осуществляется на основе подхода искусственных иммунных систем (ИИС). Формирование когнитивно-визуальных схем для динамической подачи персонализированной обучающей информации с учетом психотипа и работы со сложным современным оборудованием в лабораториях коллективного пользования повышает уровень восприятия и профессиональный уровень обучающихся.

Постановка задачи формулируется следующим образом: необходимо разработать инновационную когнитивную Smart-технология персонализированного ДО инженерным специальностям людей с ослабленным зрением в лабораториях коллективного пользования (ЛКП) на оборудовании фирмы Honeywell, которое широко применяется в нефтегазовой отрасли.

При решении поставленной задачи для определения остроты зрения обучающихся используется таблица Сивцева [19]. Для определения психотипа используется специально разработанный тест А. Белова [20], который состоит из четырех блоков вопросов и определенной формулы для расчета психотипа человека. Психотип человека является определяющей характеристикой личности и важной отправной точкой для понимания индивидуальных различий в людях [21]. Как показано в таблице, рассматриваются четыре психотипа человека: сангвиник, холерик, флегматик и меланхолик.

Таблица описания психотипов человека

Психотип	Описание психотипа	
	Сильные стороны	Слабые стороны
Сангвиник	Полон энтузиазма, щедрый, сострадательный, дружелюбный, общительный, разговорчивый, беззаботный, привлекательная личность.	Любит похвастаться, эгоист, неорганизованный, ненадежный, слабовольный, неуравновешенный, недисциплинированный.
Холерик	Активный, напористый, уверенный, мужественный, оптимист, прагматик, решительный, творческий.	Гордый, самоуверенный, невнимательный к людям, раздражительный, вспыльчивый, бесстрастный, властный.

Флегматик	Знает свое дело, располагающий, организованный, ответственный, консерватор, покладистый, надежный, спокойный.	Закрытый, эгоистичный, тяжелый на подъем, скупой, нерешительный, пассивный, робкий.
Меланхолик	Эстет, трудолюбивый, идеалист, одаренный, самоотверженный, верный, дисциплинированный.	Эгоистичный, мстительный, требовательный, пессимист, не практичный, строгий, обидчивый.

На рисунке приведена структурная схема когнитивной Smart – технологии ДО для людей с ослабленным зрением.

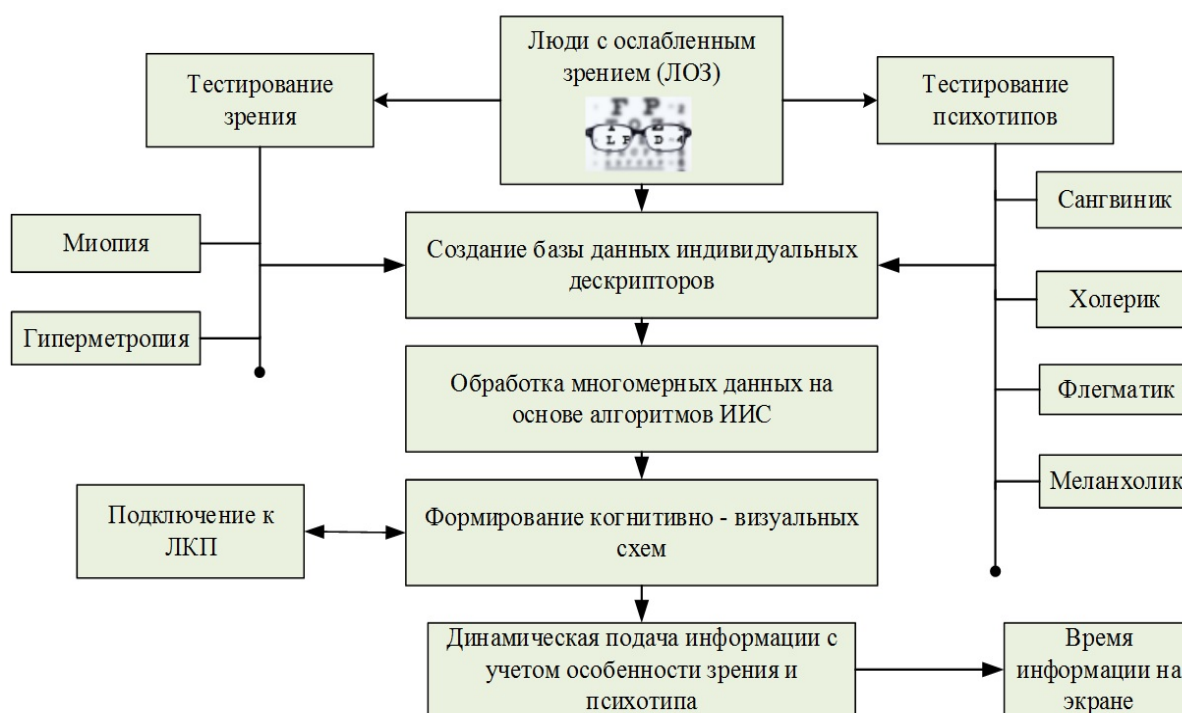


Рис. Когнитивная SMART- технология ДО для ЛОЗ.

Разработан следующий алгоритм для когнитивной Smart - технологии ДО ЛОЗ.

Люди с ослабленным зрением проходят тестирование по таблице Сивцева для определения остроты зрения. Определяется психотип обучающегося на основе тестирования по методу А. Белова. Создается база данных индивидуальных признаков (дескрипторов) для каждого обучающегося. Осуществляется предварительная обработка дескрипторов и прогноз результатов обучения на основе подхода ИИС. Далее формируются когнитивно-визуальные схемы. После подключения к лаборатории

коллективного пользования осуществляется динамическая подача учебной информации с учётом особенности зрения и психотипа обучающегося.

Список литературы

1. [Popova O.](#), [Popov B.](#), [Karandey V.](#) Intelligence Amplification in Distance Learning through the Binary Tree of Question-answer / O. [Popova](#), B. [Popov](#), V. [Karandey](#) // J. Procedia - Social and Behavioral Sciences: Elsevier. — 2015. — Vol. 214. — P. 75 – 85.
2. Yousif J.H., Saini D.K. and Uraibi H.S. Artificial Intelligence in E-learning-Pedagogical and Cognitive Aspects / J.H. Yousif, D.K. Saini, H.S. Uraibi // Proc. of the World Congress on Engineering. — London, U.K. — 2011. — Vol 2. — P.452-457.
3. Самигулина Г.А., Самигулина З.И. Разработка инновационной когнитивной SMART – технологии дистанционного обучения современному промышленному оборудованию в рамках концепции цифроватизации промышленности Казахстана / Г.А. Самигулина, З.И. Самигулина // Материалы круглого стола «Цифроватизация промышленности - основа четвертой промышленной революции». — Алматы. — 2018. — С. 88-91.
4. Bhowmick A., Hazarika S.M. An insight into assistive technology for the visually impaired and blind people: state-of-the-art and future trends / A. Bhowmick. S.M. Hazarika // [Journal on Multimodal User Interfaces](#). — 2017. — Vol. 11, [Issue 2](#). — P.149–172.
5. Шилкин Г.А. Физиологическая оптика. Рефракция, аккомодация: учебное пособие для студентов мед. ВУЗов. / Г.А. Шилкин. — Москва: Медицина. — 2007. — С.20-45.
6. Samigulina G. A., Shayakhmetova A. S., Nyussupov A. Innovative intelligent technology of distance learning for visually impaired people / G. A. Samigulina, A. S. Shayakhmetova, A. Nyussupov // J. Open engineering. — 2017. — №7. — P. 444-452.
7. Panchanathan S., Moore M., Venkateswara H., Chakraborty S., McDaniel T. Chapter 8 - Computer Vision for Augmentative and Alternative Communication / S. Panchanathan, M. Moore, H. Venkateswara, S. Chakraborty, T. McDaniel // Computer Vision and Pattern Recognition. — Academic Press. — 2018. — P. 211-248.
8. Florea C., Florea L., Vertan C. Chapter 2 - Computer Vision for Cognition: An Eye Focused Perspective / C. Florea, L. Florea, C. Vertan // Computer Vision and Pattern Recognition. — Academic Press. — 2018. — P.51-74.
9. Chin-Ling L., Chih-Hui H., Cheng-Jian L. Test-Sheet Composition Using Immune Algorithm for E-Learning Application / L. Chin-Ling, H. Chih-Hui, L. Cheng-Jian // New Trends in Applied Artificial Intelligence. — Berlin: Springer. — 2007. — P. 823-833.
10. Samigulina G.A., Samigulina Z.I. Intelligent System of Distance Education of Engineers, based on Modern innovative Technologies / G.A. Samigulina, Z.I. Samigulina // Proc. of the II International Conference on Higher Education Advances, HEAd'16. J. Procedia- Social and Behavioral Sciences. — Valencia, Spain: Elsevier. — 2016. — № 228. — P. 229-236.
11. [Milman N. B.](#) Distance Education / N. B. [Milman](#) // [International Encyclopedia of the Social & Behavioral Sciences \(Second Edition\)](#). — Elsevier. — 2015. — P. 567-570.

12. Mathewson K.J., Tang A., Fortier P., Miskovic V., Schmidt L.A. Individual Differences in Temperament: Definition, Measurement, and Outcomes / K.J. Mathewson, A. Tang, P. Fortier, V. Miskovic, L.A. Schmidt // Reference Module in Neuroscience and Biobehavioral Psychology. — Elsevier. — 2017. — P. 418-425.
13. Vasilenko E.A., Meshcheryakova T.V., Koltsova E.M., Dikaya E.A. The development of information-education materials for systems of distance learning: «Protection of Intellectual Property», an example of a course of study / E.A. Vasilenko, T.V. Meshcheryakova, E.M. Koltsova, E.A. Dikaya // [Scientific and Technical Information Processing](#). — 2011. — № 38(3). — P.193-200.
14. Seyma Kucukozer-Cavdara, Tugba Taskaya-Temizela. Analyzing the effects of the personality traits on the success of online collaborative groups / Seyma Kucukozer-Cavdara, Tugba Taskaya-Temizela // Proc. of 2nd International Conference on Higher Education Advances. — València, Spain, Elsevier. — 2016. — P. 37-923.
15. Coulacoglou C., Donald H. Saklofske. Chapter 7 - The Role of Temperament in Development and Psychopathology / C. Coulacoglou, H. Donald Saklofske // Psychometrics and Psychological Assessment. Principles and Applications. — Academic Press. — 2018 — P. 157-185.
16. Rothbart M.K. Temperament and Human Development / M.K. Rothbart // International Encyclopedia of the Social & Behavioral Sciences (Second Edition). — Elsevier. — 2015. - P.180-183.
17. Mathewson K.J., Tang A., Fortier P., Miskovic V., Schmidt L.A. Individual Differences in Temperament: Definition, Measurement, and Outcomes / K.J. Mathewson, A. Tang, P. Fortier, V. Miskovic, L.A. Schmidt // Reference Module in Neuroscience and Biobehavioral Psychology. — Elsevier. — 2017. — P. 418-425.
18. Xiaoxue Fu, Koraly Pérez-Edgar. Temperament Development, Theories of / F. Xiaoxue, P.E. Koraly // International Encyclopedia of the Social & Behavioral Sciences (Second Edition). — Elsevier. — 2015. - P. 191-198.
19. Копаева В.Г. Глазные болезни / В.Г. Копаева // Основы офтальмологии: учебник. — Москва: Медицина. — 2012. — С.510-560.
20. Тест на темперамент А. Белова. Формула темперамента. E-print: <https://psycabi.net/testy/202-test-na-temperament-a-belova> (11.07.2018).
21. Samuel P. Putnam. Temperament / P. Samuel Putnam // International Encyclopedia of the Social & Behavioral Sciences (Second Edition). — Elsevier. — 2015. — P. 173-179.

ЗАПОЛНЕНИЕ ПУСТОТ СРЕДСТВАМИ ДИС-ТЕХНОЛОГИИ

В.П.Сизиков¹, В.И. Разумов²

¹Омский государственный университет путей сообщения, v_p_sizikov@mail.ru

²Омский государственный университет им. Ф.М. Достоевского, razumovvi@omsu.ru

1. Введение. Тематика заполнения пустот, в принципе, неизбежна при использовании любой технологии моделирования, так как последняя имеет дело с дискретными данными, да и по ним часто нет гарантии о надёжности и полноте. Будь то проблема распознавания или идентификации, вопрос о начальных или граничных данных при постановках задач, каждый раз приходится обдумывать и оценивать, как скажется на результатах, на степени их адекватности и точности, отсутствие сведений в иных местах, отличных от избранных дискретных.

Естественно, указанная тематика является одной из главных составляющих работы по использованию моделей в ранге нейросетей, к примеру, [1]. Не избавлено, конечно, от этого и моделирование средствами ДИС-технологии, тем более что имеется взаимная связь между моделями в ранге динамических информационных систем (ДИС) и между моделями в ранге нейросетей [2].

Однако существенной отличительной особенностью ДИС-технологии является факт наличия при ней файла с номологической базой (НБ) [3], расположенного по адресу [<http://thoughtring.com/ViewForm.aspx?id=421>], что позволяет автоматизировать [4] формирование качественных моделей объектов и процессов в ранге ДИС-компьютеров (ДИС-*К). По сути, модель здесь сразу оказывается вполне определённой на уровне НБ, а факт заполнения выступает всего лишь интерпретацией категорий НБ в терминах избранной предметной области и соответствующим осмыслением этих терминов. Если в традиции преобладает движение от предметной области к модели без гарантии обеспечения системности, то ДИС-технология позволяет двигаться больше, наоборот, от модели к предметной области, дорабатывая по ходу системные моменты. При посредстве НБ для конструирования качественных моделей, ответственных за передачу содержательной специфики объекта, ДИС-технология сохраняет их способности к однозначности и точности выражения, свойственной алгоритмическим моделям.

Как-никак, в традиции преобладает подход, при котором организация чего-то имеет в основе постулат практически неограниченного служения её во времени и полной её ликвидации в случае, как только потребуется некая в ней перемена. А там, где перемены наступают скоротечно, предпочитают вообще не мечтать об организации и полагаться на случайные разовые процедуры, контроль за которыми отсутствует. Оба этих варианта ведут к конфликтам на социальном уровне, к потере доверия, морали, ответственности, совести и многих других ценных качеств человека. Необходимо выстраивать организации, сразу заботясь об их готовности к новым переменам. Для этого нужна смысловая системная проработка организации, что достигается пока лишь в рамках ДИС-технологии [3].

Приводится образец на примере инженерии программирования. В принципе, тема выбрана совершенно случайно, так что по аналогии с изложенным могут быть осуществлены проработки и широкого многообразия других тем. Главное, чтобы на нужную тему имелась информация в постепенно пополняемом файле НБ [3].

Разумеется, наряду с проработкой качественных моделей в ДИС-технологии приходится иметь дело и с алгоритмическими моделями, в которых внимание отводится функциональному аспекту систем и требуется определяться с функциональными параметрами ДИС. Однако и здесь рабочая дорога предстаёт пока как однонаправленная, экспериментальная, так как традиционные представления об объектах и процессах не являются системными и извлекать в их рамках данные, которые можно было бы

анализировать на системном уровне, определять по ним значения функциональных параметров нужной ДИС, практически безнадежно. Необходимы переход от традиционной вещественно-энергетической парадигмы к информационно-полевой и выработка в рамках последней качественно новых представлений о природе объектов [5].

2. Формирование развёртки подходящего универсального понятия. Чем полнее и детальнее развернём нужное понятие, тем лучше. Поэтому, заглядывая в файл НБ, отдаём предпочтение искать имя, близкое по смыслу к «инженерии программирования», в списке ДИС-*К уровня 3. И среди таких имён находится просто «инженерия».

Понятие «инженерия» в НБ представлено 27-вершинной кодировкой (037ABGKMO)(0589CDJLQ)(157EFHINP). Пытаемся выделить в каждом из блоков этой кодировки по три индекса таких, чтобы образовавшиеся из них блоки давали уже известную в НБ 9-вершинную кодировку определённого понятия. Отталкиваясь от начального индекса кодировки, обнаруживаем, что он оказывается таковым в 9-вершинной кодировке (07B)(05D)(57F), выражающей понятие «идентифицированное понимание». Повторяем далее такую процедуру с оставшимися после этого индексами исходной 27-вершинной кодировки, но уже с дополнительным условием, что оставшиеся после неё индексы тоже должны входить в некую известную 9-вершинную кодировку. Если такого добиться не удастся, значит, первоначально выбранная 9-вершинная кодировка оказалась нехотой, и её следует заменить другой. К счастью, удастся выбрать 9-вершинную кодировку (3AO)(8CJ)(1HN), выражающую понятие «проектирование», после которой остаётся 9-вершинная кодировка (GKM)(9LQ)(EIP), выражающая понятие «проектная связь». Так понятие «инженерия» представляет триаду как ДИС-*К D₁ уровня 1 из понятий «идентифицированное понимание», «проектирование», «проектная связь».

Теперь поступаем аналогично с каждым из выявленных понятий, выявляя в их 9-вершинных кодировках 3-вершинные части, выбирая в каждом из блоков кодировки по одному индексу. При этом лучше ориентироваться на выбор таких 3-вершинных частей, что помечены в НБ жирным шрифтом и имеют статус законов [3], согласующих знания и реальность. В итоге получим, что понятие «идентифицированное понимание» представляет триаду из понятий: «гносеологическая ёмкость знания» с кодировкой (0)(5)(7), «операционный потенциал» с кодировкой (7)(0)(5), «качество модели» с кодировкой (B)(D)(F). Аналогично, понятие «проектирование» представляет триаду из понятий: «текст» с кодировкой (3)(8)(1), «функциональность структуры» с кодировкой (A)(C)(H), «стационарная связь» с кодировкой (O)(J)(N). Наконец, понятие «проектная связь» представляет триаду из понятий: «прототип» с кодировкой (G)(9)(E), «взаимодействующая ведущая связь» с кодировкой (K)(L)(P), «взаимодействующая контролирующая связь» с кодировкой (M)(Q)(I). В результате исходное понятие с 27-вершинной кодировкой может быть представлено уже в форме ДИС-*К D₂ уровня 2 из 9 полученных понятий с 3-вершинными кодировками (рис. 1), причём эти понятия имеют в НБ статус законов.

Наконец, каждое из понятий с 3-вершинной кодировкой представляет триаду мини-категорий, имена которых прописаны в начале НБ согласно их индексам. В результате исходное понятие с 27-вершинной кодировкой будет представлено уже в форме ДИС-*К D₃ уровня 3 из 27 мини-категорий (рис. 2). Примем во внимание, что при изображении ДИС-*К D₂ не имеет значения выбор порядка у понятий на обоих этапах дешифровки, а при изображении ДИС-*К D₃ выбор порядка у понятий на последнем этапе дешифровки произвола не допускает, но определяется однозначно согласно

правилам по связям между блоками V_0 , V_1 , V_2 , под которые и отводятся блоки кодировок.

Для осмысления и надёжности обеих развёрток следует к изображению ДИС-*К D_2 добавить изображения остальных его 5 базовых мутаций [4] (рис. 1). Это потребует дополнительного осмысления возникающих новых триад с их 9-вершинными кодировками. Велика вероятность, что эти триады уже тоже представлены в НБ, так что останется лишь найти их имена. По крайней мере, заведомо есть хотя бы одна из возможных развёрток исходного понятия, при которой все возникающие триады будут представлены в НБ. Однако не следует теряться, если такого не случилось, и можно попытаться привести свои версии новых имён, не повторяющих, конечно, те, что уже есть в НБ. Более того, информация с новыми именами может быть прислана на сайт [<http://thoughttring.com/ViewForm.aspx?id=421>] и использована для пополнения НБ после соответствующей экспертизы. Аналогично поступаем и с изображением ДИС-*К D_3 , применяя уже базовые мутации по отдельности к каждой из его 9-вершинной части [4]: «идентифицированное понимание», «проектирование», «проектная связь» (рис. 2). Из соображений удобства здесь можно изображать не ДИС-*К уровня 3, а ДИС-*К уровня 2 для каждой части в отдельности, однако при этом их количество увеличится вдвое.

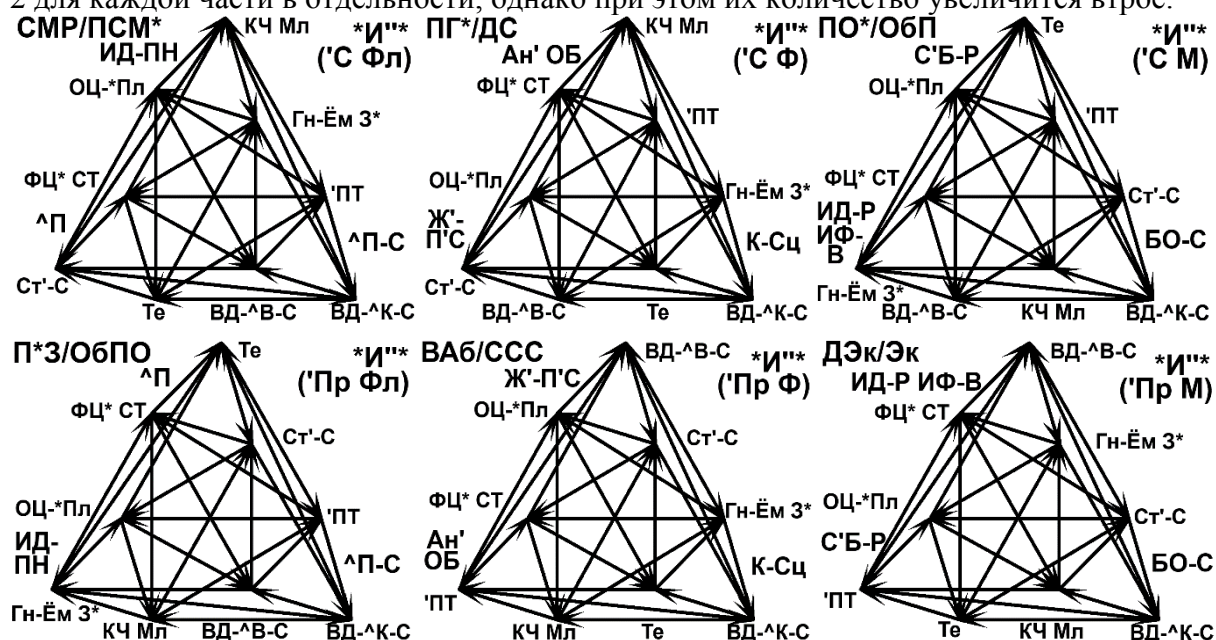


Рис. 1. Базовые мутации схемы инженерии (*I'') на уровне законов номологической базы. Обозначения: Ан' – анимация; БО-С – безопасная связь; ВАБ – воплощение абстрактного; ВД-^В-С – взаимодействующая ведущая связь; ВД-^К-С – взаимодействующая контролирующая связь; Гн-Ём – гносеологическая ёмкость; ДС – давление страха; ДЭк – доступ к эксперименту; Ж'-П'С – жанровая постановка; 3* – знание; ИД-ПН – идентифицированное понимание; ИД-Р – идентифицирующая роль; ИФ-В – информационная величина; К-Сц – когнитивный сценарий; КЧ – качество; М – математика; Мл – модель; ОБ – объект; ОБП – обучение подражанием; ОБПО – обучение практическим освоением; ОЦ-*Пл – операционный потенциал; ^П – проектирование; ПГ* – проверка гипотезы; П*3 – понимание закономерностей; ПО* – понимание опыта; 'Пр – приложение; ^П-С – проектная связь; ПСМ* – подключение к саморазвитию Мироздания; 'ПТ – прототип; 'С – становление; С'Б-Р – симбиотическая роль; СМР – саморазвитие; ССС – созидание средств страхования; СТ – структура; СТ'-С – стационарная связь; Те – текст; Ф – физика; Фл – философия; ФЦ* – функциональность; Эк – эксперимент.

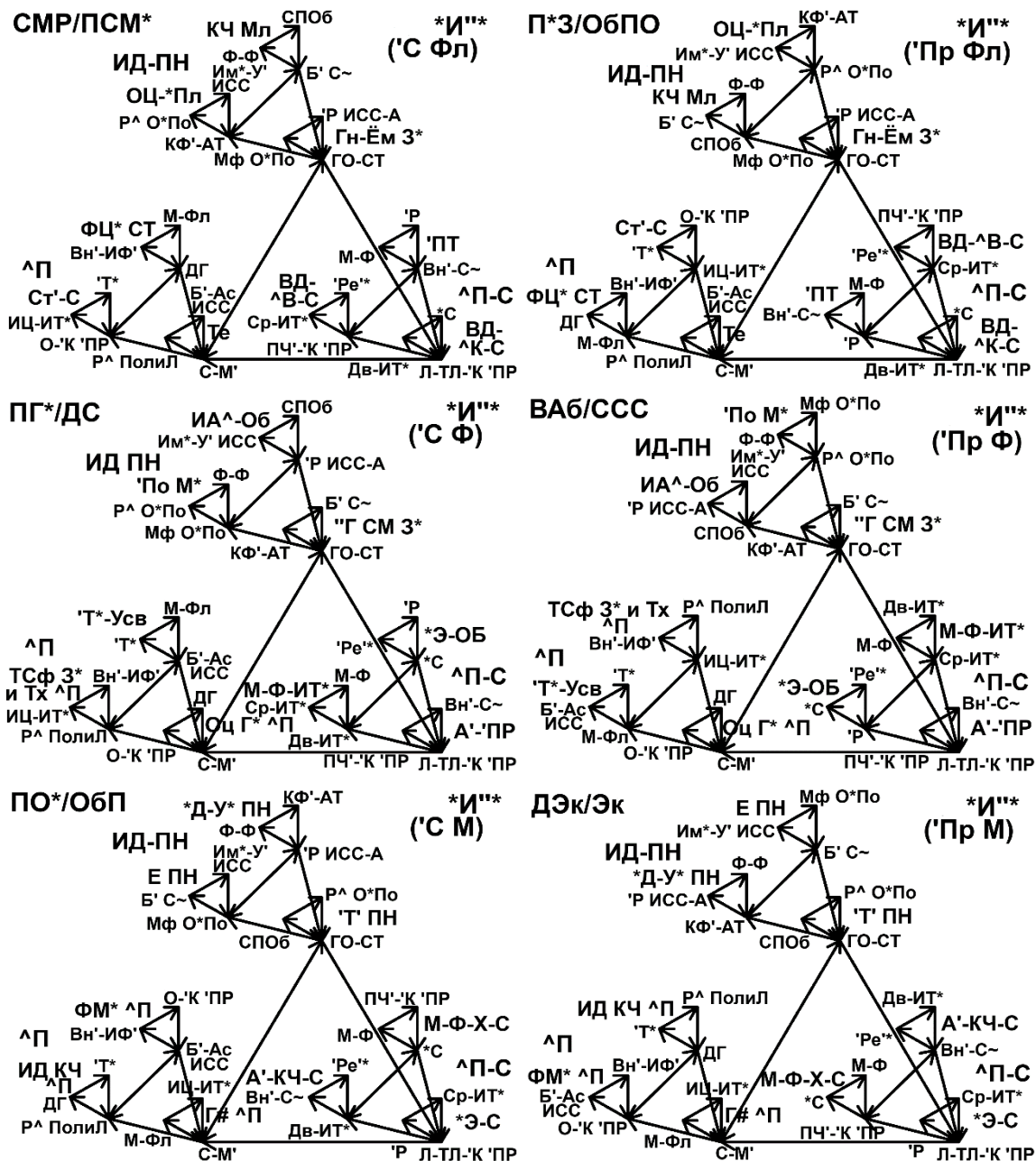


Рис. 2. Базовые мутации схемы инженерии (*И''') на уровне мини-категорий номологической базы. Обозначения, дополняющие таковые с рис. 1: А'-КЧ-С – адаптивная качественная связь; А'-ПР – адаптивное принятие решений; Б' – база; Б'-Ас – базовый аспект; Вн'-ИФ' – внешний интерфейс; Вн'-С~ – внешний синтез; 'Г – генерация; Г# – герменевтика; Г* – гомеостаз; ГО-СТ – генетически обусловленная структура; Дв-ИТ* – интерпретации на языке движения; ДГ – диагностика; *Д-У* – динамическая устойчивость; Е – единство; ИА^ -Об – интерактивное обучение; ИД – идентификация; Им*-У' – имитационный уровень; ИСС – исследование; ИСС-А – исследовательский аппарат; ИЦ-ИТ* – интерпретации на языке инициирования; КФ'-АТ – конфигурационный автомат; Л-ТЛ-К – логико-топологический компонент; М* – Мироздание; Мф – метафизика; М-Ф – математическая физика; М-Ф-ИТ* – математико-физические интерпретации; М-Фл – математическая философия; М-Ф-Х-С – математико-физическая характеристическая связь; О-К – онтологический компонент; О*По – оппозиция; Оц – оценка; ПН – понимание; 'По – полярность; ПолиЛ – полилогичность; 'ПР – принятие решений; ПЧ'-К – причинный компонент; Р^ –

развёртка; 'Р – развитие; 'Ре'* – репрезентативность; С~ – синтез; *С – сущность; СМ – система; С-М' – связная мутация; СПОб – состоятельный процесс обучения; Ср-ИТ* – интерпретации на языке среды; 'Т' – трансляция; 'Т*' – трансформатор; ТСф – трансфер; 'Т*-Усв – трансформаторное усвоение; Тх – технология; ФМ* – фундаментальность; Ф-Ф – физическая физика; *Э-ОБ – эволюционирующий объект; *Э-С – эволюционная связь.

3. Формирование развёртки требуемого понятия. Теперь предстоит провести интерпретацию возникших при развёртывании понятия «инженерия» категорий НБ под его частный случай в ранге «инженерии программирования». Эта процедура тяжелее предыдущей, так как требует формирования, по сути, новых категорий на базе представлений из жизненного опыта по теме программирования. Интерпретация схем с рис. 1 даётся, конечно, проще не только потому, что в них меньше категорий, но и за счёт ассоциаций этих категорий с законами реальности. Кроме того, содержание этих категорий при необходимости может быть дополнительно прояснено за счёт обращения к их дешифровкам на рис. 2.

Как пример для «инженерии программирования» можно выбрать следующие интерпретации схем на рис. 1. Идентифицированное понимание – потенциал программирования: гносеологическая ёмкость знания – знания по программированию; операционный потенциал – набор операций в программировании; качество модели – освоенная модель программирования. Проектирование – организация программирования: текст – описание сведений по программированию; функциональность структуры – насколько структура программирования поддерживает его функциональность; стационарная связь – инициативы по осуществлению программирования. Проектная связь – выходы программирования на эксплуатацию модели: прототип – место программирования в деле использования модели; взаимодействующая ведущая связь – ориентации на выносимость модели; взаимодействующая контролирующая связь – ориентации на поведение модели. Когнитивный сценарий – сценарий поведения модели; жанровая постановка – оперативная процедура программирования; анимация объекта – эффективность программирования; идентифицирующая роль информационной величины – готовность программирования к обновлению функций; симбиотическая роль – готовность программирования к модернизации; безопасная связь – ориентации на безопасность эксплуатации модели.

В результате получаем конкретный набор базовых схем по инженерии программирования, аналогичных схемам с рис. 1. Этот набор можно расценить как законодательную базу для осуществления инженерии программирования в статусе системы, для обеспечения перемен в программировании, минуя обречения. Этой базы уже достаточно для устойчивого поддержания организации программирования.

Что касается схем с рис. 2, то особой необходимости для осуществления их интерпретации нет. Однако к прописанным в них детализациям актуально будет обращаться в текущем режиме для уточнения или корректировки дел по программированию. Кроме того, детализации могут оказаться востребованы для категорий не первой схемы с рис. 1, а второй или третьей, для которых на рис. 2 нет развёрток. Так, вполне вероятно, что может понадобиться внесение изменений в оперативную процедуру программирования, и тогда важно будет детальнее развернуть и осмыслить блок «жанровая постановка» с рис. 1, чего нет непосредственно на рис. 2. Однако ориентиры для этого на рис. 2 всё же есть в ранге дешифровок до мини-категорий у законов НБ, из которых складывается блок «жанровая постановка». И таких ориентиров тоже может оказаться вполне достаточно для решения текущих проблем с изменениями

в оперативной процедуре программирования. В частности, возможно, придётся вести программирование, исходя из другой базы данных.

4. Заключение. Таким образом, изложена тематика заполнения пустот на примере методики смысловой системной проработки с использованием НБ [3], сформированной в рамках ДИС-технологии. Это актуально и весьма необходимо для всего, что работает или используется в статусе системы, так как позволяет предусмотреть качества адаптации и готовности к переменам, минуя обречения.

На настоящее время тема подбора отсутствующих данных на стадии построения качественной модели объекта с соблюдением требований алгоритмической строгости развития не получила. ДИС-технология, развернутая на НБ, позволяет сделать дополнительными качественный (содержательный) и алгоритмический подходы к моделированию.

Тема инженерии программирования выбрана исключительно в качестве примера, так что по аналогии с изложенным могут быть осуществлены проработки и широкого многообразия других тем – воля, коммуникация, математический анализ, мультипликатор, редукция, цель. Главное, чтобы на нужную тему имелась информация в постепенно пополняемом файле НБ [<http://thoughttring.com/ViewForm.aspx?id=421>].

Несмотря на произвольность выбора темы для изложений методики, осуществлённая по инженерии программирования проработка может быть использована как ориентир по делу устойчивой организации программирования, адаптивной к разнообразным переменам. В частности, эта проработка чётко указывает, что не следует замыкаться на какой-то одной технологии программирования, всё подгонять под неё, а потом переживать, что такая технология вдруг утратила значение и для выхода на новое надо «горы сворачивать», на что ни сил, ни средств не хватает.

Конечно, немалая польза была бы и от проработки функционального аспекта у сформированной модели. Однако, как отмечено вначале, здесь возможен пока лишь вариант экспериментов на подбор значений желаемых функциональных параметров с приличной в этом помощью со стороны известных специальных теорем. Вместе с тем, это указывает на актуальность перехода от традиционной вещественно-энергетической парадигмы к информационно-полевой и выработку в рамках последней качественно новых представлений о природе объектов [5].

Список литературы

1. Глазков А.В. Моделирование процессов распознавания и классификации образов на основе нейросетей / А.В. Глазков, А.В. Крутов // Вестник Воронежского государственного технического университета. – 2009. – Т. 5. – № 11. – С. 240–246.
2. Разумов В.И. Основы теории динамических информационных систем / В.И. Разумов, В.П. Сизиков. – Омск: ОмГУ, 2005. – 212 с.; То же [Электронный ресурс]. Режим доступа: <http://www.omsu.ru/file.php?id=4264>.
3. Разумов В.И. Приложения информационных основ синтеза систем: ДИС-развёртки и парадигма закона. Монография [Электронный ресурс] / В.И. Разумов, В.П. Сизиков. Электронное издание № 33934. № гос. регистр. 0321304636, ФГУП НТИЦ «Информрегистр». – Омск. 2014.
4. Разумов В.И. Автоматизация рассуждений: программирование мутаций ДИС-компьютера уровня 2 / В.И. Разумов, В.П. Сизиков // Сибирский философский журнал. – 2016. – Т. 14. – № 1. – С. 53–69.
5. Разумов В.И. К новой парадигме закона / В.И. Разумов, В.П. Сизиков // Вестник Омского ун-та. – 2012. – № 2. С. 213–219.

ОБРАБОТКА БАЗЫ ЗНАНИЙ В СИСТЕМЕ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ИНСТРУКТОРОВ ПРИ ВЫБОРЕ ПРОГРАММ ТРЕНИРОВКИ В ЛЮБИТЕЛЬСКОМ СПОРТЕ

Травкина Е.С.^{1,2}, В.А.Углев², Ступницкий М.С.³

¹Спортивно-танцевальная студия «Пурпур», katya_ml15@mail.ru

²Сибирский федеральный университет, uglev-v@yandex.ru

³ФГБУЗ КБ №51 ФМБА России, stmishgun@gmail.com

В связи с тем, что в последние годы рынок предложений в сфере любительского спорта существенно возрос, любительский непрофессиональный спорт стал все более доступен для человека. Под любительским непрофессиональным спортом в данной статье следует понимать такие направления тренировок, которые позволяют совершенствовать физические данные человека без применения специализированного оборудования (велотренажёры, брусья, штанга и т.д.). Такие тренировки могут проводиться как с тренером, так и самостоятельно, в спортивном зале или в домашних условиях.

Рост популярности любительского спорта привел к обострению тех проблем, которые традиционно связаны с принятием решений при выборе упражнений. С одной стороны, с ростом числа представителей профессии тренер/инструктор и доступности обучения этой профессии растет число недостаточно квалифицированных кадров, с ограниченными знаниями в области медицины. С другой стороны, любительский спорт привлекает многие категории населения, среди которых часто встречаются люди с проблемами со здоровьем. Таким людям физическая нагрузка может быть необходима, но при этом важно качественно оценивать состояние здоровья перед началом занятий любительским непрофессиональным спортом, во избежание нанести урон здоровью человека. Однако, решение проблемы компетентной оценки состояния здоровья спортсменов-любителей⁴ при формировании программы тренировок не является тривиальным. Достаточно сложно укомплектовать все спортивные центры медицинским персоналом, для обеспечения контроля самочувствия спортсменов и для предоставления услуги по формированию программ тренировок, так как это приведет к росту цен на все услуги в таких центрах и спровоцирует отток спортсменов. Поэтому тренера, в случае отсутствия медицинского образования, неизбежно сталкиваются с проблемой комплексной оценки состояния здоровья спортсмена и его физического развития [1] при формировании программы упражнений для спортсменов-любителей. Под комплексной оценкой понимается учёт заболеваний у спортсменов, определение противопоказаний к определенным типам занятий или упражнениям, оценка физической подготовки спортсмена любителя по различным критериям (например, сила, выносливость, гибкость и пр.).

У описанной проблемы в настоящее время уже имеются несколько решений. Одними из них является личная беседа с профессиональным тренером (он же эксперт в области физической культуры и спорта) в спортивном зале. Эксперт составляет индивидуальный

⁴ Для краткости, далее будет использоваться термин спортсмен.

перечень упражнений с учетом пожеланий спортсмена-любителя. При принятии решения о включении/не включении упражнений в перечень, эксперт руководствуется собранной информацией от спортсмена: о его данных, параметрах тела, проблемах со здоровьем, о целях и возможностях заниматься на регулярной основе. На основании полученной информации эксперт составляет перечень упражнений, который приведет спортсмена к желаемому результату. Одними из недостатков этого подхода являются: высокая цена за консультацию и составление перечня упражнений; при отсутствии у эксперта специальных медицинских знаний, повышается риск заполучить травму или ухудшить состояние здоровья. При невозможности обращения к эксперту за помощью в составлении перечня упражнений можно воспользоваться онлайн-сервисами по составлению программ тренировок для тренажерного зала или для занятий дома. Данные сервисы функционируют по следующему принципу: требуется заполнить анкету, в которой указываются личные данные, желаемое количество тренировок в неделю, цель тренировок, тип тренировок, вопросы о физическом состоянии. На основе полученной информации сервис выбирает из множества существующих в репозитории программ наиболее подходящую программу тренировок для конкретного пользователя. Недостатками при таком подходе являются: отсутствие подробной оценки состояния здоровья пользователя; сложность выявления составителя перечней упражнений, для того чтобы убедиться в его компетентности и образованности в данной области.

Не так давно появилась разработка в области автоматизированного составления программ тренировок – система «WellnessSystem» от итальянской компании TechnoGym. Данный программный продукт используется только в работе спортивных центров, оборудованных тренажерами данного производителя. Каждый новый посетитель спортивного центра указывает возраст, рост, вес, противопоказания, заболевания. На основании собранной информации система предлагает готовую программу тренировок, наиболее подходящую из всего множества программ, имеющихся в памяти. При этом посетитель имеет возможность составить свою программу тренировок из предложенных упражнений, либо воспользоваться помощью тренера. Данный подход имеет свои минусы: услуги спортивных центров, оснащенных системой «WellnessSystem» имеют высокие цены, и при этом данная система не имеет широкого распространения; перечни упражнений выбираются автоматически из готового числа программ тренировок, либо формируются тренером; существует риск составления не эффективного перечня упражнений, который будет либо не отвечать целям спортсмена, либо не в полной мере учитывать особенности здоровья спортсмена.

Указанные недостатки существующих подходов предлагается решить путем автоматизированного синтеза перечней упражнений с использованием интеллектуальной системы поддержки принятия решений, основанной на базе экспертных систем с доской объявлений. При этом необходимо уйти от случая, когда выбирается один из готовых вариантов программы тренировок, отдав предпочтение синтезу индивидуализированной программы на уровне отдельных упражнений. Рассмотрим основные этапы работы с системой:

Первым этапом в работе системы является заполнение спортсменом анкет, выявляющих общие сведения, такие как возраст, рост, вес, опыт в занятиях спортом, желаемое количество тренировок в неделю и пр. Далее на выбор предоставляется шесть целей тренировок: сбросить вес, улучшить гибкость, улучшить рельеф, поддержание формы, развить силу, развить выносливость (поддерживается выбор их комбинации). На основании результатов первой анкеты и выбранной цели спортсмен может получить комментарий системы, касающийся реалистичности достижения целевого состояния за указанный период времени при текущих условиях. На следующем шаге анкетирования происходит выявление противопоказаний к определенным видам нагрузки или отдельным упражнениям: спортсмен заполняет подробную анкету о заболеваниях систем организма, о состоянии опорно-двигательного аппарата и физического самочувствия в данный момент времени. Полученные результаты будут являться основополагающими критериями при принятии решения о включении или не включении конкретных упражнений в программу. При этом допускается, что на ряд вопросов спортсмен может не дать чёткого ответа или затрудниться дать его вообще.

На *втором этапе* проводится тест на оценку физического состояния спортсмена: оцениваются силовые показатели, скоростные, гибкость и растяжка при помощи специально подобранных упражнений⁵. С помощью выявленных данных будет рассчитана допустимая интенсивность выполнения каждого упражнения в программе.

На *третьем этапе*, на основании полученных данных от спортсмена, принимается решение о подходящем типе программы, которая может быть последовательной или круговой. Для каждого типа программы определяется свой шаблон, исходя из цели и количества тренировок в неделю. Шаблон программы тренировок представляет собой перечень нагружаемых групп мышц в каждый конкретный день тренировки и количество упражнений для каждой группы⁶. Выбор упражнений происходит интеллектуальным ядром системы поддержки принятия решений.

Завершающий (четвертый) этап заключается в синтезе бланка-отчёта системы, содержащего как перечень упражнений (программу тренировок), так и комплекс рекомендаций по их выполнению, относительно индивидуальных особенностей спортсмена.

Реализация работы такой системы поддержки принятия решений предполагает проработку модуля принятия решений, реализуемого в виде экспертной системы на прецедентах [2]. Данный тип экспертных систем был выбран потому, что в нем имеется чёткое разделение поля знаний на различные профили (базы знаний *спортивного* и *медицинского* профиля), а так же итерационное принятие компромиссных решений по

⁵ Упражнения, входящие в тест, просты и могут выполняться самостоятельно, а результаты их выполнения вводиться в соответствующие диалоговые поля системы. Как правило, они имеют количественную природу и не требуют специальных познаний [1].

⁶ Программа тренировок составляется, как правило, из числа занятий за одну неделю. Но её структура сильно зависит от цели тренировок и периода, за который желательно достигнуть поставленный результат.

каждому включаемому в шаблон упражнению. Если укрупнено рассмотреть цикл работы решателя (третий этап), то он состоит из следующих шагов:

1. Исходные данные концентрируются и фазифицируются для подачи на вход в экспертную систему.
2. Выбирается тип шаблона, инициализируя доску объявлений.
3. Шаблон заполняется наиболее эффективными упражнениями, отвечающими цели спортсмена, относительно спортивной базы знаний (далее БЗ).
4. Программа тренировок сокращается за счёт тех упражнений, которые исключаются на основании противопоказаний и имеющихся заболеваний спортсмена (отрабатывает медицинская БЗ).
3. Формируется список рекомендуемых упражнений с точки зрения медицинской БЗ (с целью профилактики и лечения имеющихся заболеваний).
4. Для каждой группы мышц выбираются альтернативные упражнения, опираясь как на цель тренировок, так и на список из шага 3 (спортивная БЗ).
5. Для каждого упражнения подбирается адекватное число подходов (интенсивность) в как рамках всей программы, так и относительно отдельной тренировки (спортивная БЗ).
6. Каждое промежуточное решение согласуется с медицинской БЗ.

Таким образом, шаблон программы тренировок циклически заполняется упражнениями, а базы знаний активизируются поочередно (шаги 4-6) до тех пор, пока общее решение не будет найдено.

Рассмотрим основные компоненты, формирующие интеллектуальное ядро системы поддержки принятия решений, а так же связи между ними:

- решатель – алгоритм, обеспечивающий просчёт экспертных систем и вывод объяснений;
- планировщик – алгоритм, координирующий вызов баз знаний и работу с общим пространством решений (доска объявлений);
- набор баз знаний, обеспечивающих поиск решений в предметной области;
- набор шаблонов различных программ тренировок.

Реализация проекта системы базируется на схеме взаимодействия компонентов, связанных между собой по модульному принципу: внешние файлы экспертных систем, модуль решателя, модуль планировщика и интерфейсный блок. В разрабатываемом прототипе интеллектуальной системы поддержки принятия решений для связывания баз знаний с проектом используется механизм динамически подключаемых библиотек (DLL) [4]. Преимуществом данного подхода является возможность независимой

загрузки базы знаний в программу FLM_Builder [5]⁷ и модификация её логики работы без перекомпиляции проекта, который её использует. Так же следует отметить, что для предобработки количественных входных данных применяется механизм нечёткой логики [6], нацеленный на сокращение размерности пространства правил в каждой базе знаний системы поддержки принятия решений.

В заключении хотелось бы отметить, что разрабатываемый прототип интеллектуальной системы поддержки принятия решений для инструкторов может быть реализован не только в качестве приложения для персонального компьютера, но в качестве интернет-сервиса. Это позволит расширить круг потребителей, предпочитающих не обращаться в спортивные центры. Кроме того, функциональность системы можно расширить за счёт учёта динамики показателей физического состояния спортсмена-любителя при синтезе дальнейших рекомендаций к тренировкам.

Список литературы

1. Дубровский В.И. Спортивная медицина. – М.: ВЛАДОС, 2002. – 512 с.
2. Джексон П. Введение в экспертные системы. – М.: Вильямс, 2001. – 624 с.
3. Гаврилова Т.А., Кудрявцев Д.В, Муромцев Д.И. Инженерия знаний. Модели и методы. – СПб.: Лань, 2018. – 324 с.
4. Углев В.А., Добронев Б.С. Разработка экспертных систем с применением внешних модулей // Молодёжь и наука: начало XXI века: Материалы Всероссийской научно-технической конференции студентов, аспирантов и молодых учёных: В 3 ч. Ч. 1. – Красноярск: ИПЦ КГТУ, 2006. – С. 305-306.
5. Углев В.А. Модуль создания экспертных систем с нечёткой логикой для среды программирования Delphi «FLM_modul.pas». – М.: ОФАП, 2005. – номер госрегистрации (ЕСПД) – 03524577.01145-01 99 01.
6. Заде Л. Понятие лингвистической переменной и его применение к принятию приближенных решений: Пер. с англ. Математика. Выпуск 3. – М.: Мир, 1976. – 168 с.

⁷ Не смотря на то, что программа FLM_Builder оперирует продукционными моделями экспертных систем, их комбинация и внешнее управление последовательностью вызова решателя для просчёта различных баз знаний (блок планировщика) позволяют реализовать практически произвольную логику вывода знаний.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ОРГАНИЗАЦИИ ДАННЫХ РЕЕСТРА ФРАГМЕНТИРОВАННЫХ ЭЛЕКТРОННЫХ РЕСУРСОВ

В.А.Углев

Сибирский федеральный университет, uglev-v@yandex.ru

Анализ результатов интеллектуальной деятельности в университетах и научных организациях во многом сводится к обработке статистики публикаций и обращения к ним (цитируемости). Традиционно, цитируемость рассчитывается для *объектов интеллектуальной собственности* (ОИС), имеющих символическую природу (тексты статей, монографий, книг и т.п.). Не смотря на это, можно выделить множество прочих типов ОИС, которые потенциально могут быть депонированы (внесены в электронный реестр)⁸. К ним могут быть отнесены САД-модели для станков с ЧПУ, массивы точек отсканированного в 3D реального объекта, последовательности нуклеотидов (фрагменты секвенированного генома), обучающие выборки (коллекции примеров) для алгоритмов машинного обучения, онтологии, электронные образовательные ресурсы и многое другое. Большинство из них вообще не подпадают под защиту закона об авторском праве (см. [1]). Более того, ОИС может иметь не только фрагментированность контента, но и неоднородность авторского вклада. По этой причине представляет интерес рассмотрение проекта создаваемого реестра [2], позволяющего обеспечить депонирование, хранение и, как следствие, оценку цитируемости подобных объектов. В качестве базового механизма представления будем опираться на язык концептуального описания SysML [3].

Перечислим те требования к реестру, которые важны с точки зрения оценки цитируемости: возможность описывать составные (фрагментированные) ОИС, возможность учитывать различный вклад соавторов в части ОИС, возможность фиксировать факты обращений и заимствований через технологию распределенных реестров (BlockChain). Учитывая основные бизнес-процессы и интересы заинтересованных сторон, построим укрупненную *концептуальную модель предметной области* (рис. 1). Фрагментированность и неоднородность авторского вклада на рисунке скрыты за блоком ОИС и требуют ясности в терминологии. Поэтому введём понятие **фрагментированный электронный ресурс** (ФЭР) – это электронный ресурс, который имеет неоднородную (фрагментированную) структуру, определяемую, как правило, семантикой объекта, типом включенных данных и/или правовым статусом фрагментов. Очевидно, что ФЭР состоит из собственно депонируемого ОИС (файла или их набора) и метаданных, детально его описывающих.

Структура метаданных ФЭР представлена в виде *модели декомпозиции* ОИС как фрагментированной сущности (рис. 2). На рисунке выделено 4 уровня иерархии детализации структуры ФЭР, а так же слой показателей качества (здесь N – порядковый номер версии ОИС, m – идентификатор фрагмента в дереве иерархии, k – число структурных элементов нижнего уровня). Охарактеризуем кратко каждый из уровней.

Уровень представления – совокупность метаданных, позволяющих идентифицировать объект ОИС в реестре и описать его инвариантные параметры.

⁸ Далее будут обсуждаться только те типы ОИС, которые депонируются в электронном виде.

Включает ключ, заглавие, внешний идентификатор IPC⁹, дату депонирования первой версии, число версий, краткую аннотацию на 1 или более языках и пр.

Уровень экземпляров – множество экземпляров метаданных, каждый из которых содержит: идентификатор уровня представления, дату депонирования, множество авторов, тип лицензии, множество ключевых слов, массив ссылок на фрагменты и их описание, статистику обращений, множество рецензий, множество комментариев/отзывов и пр.

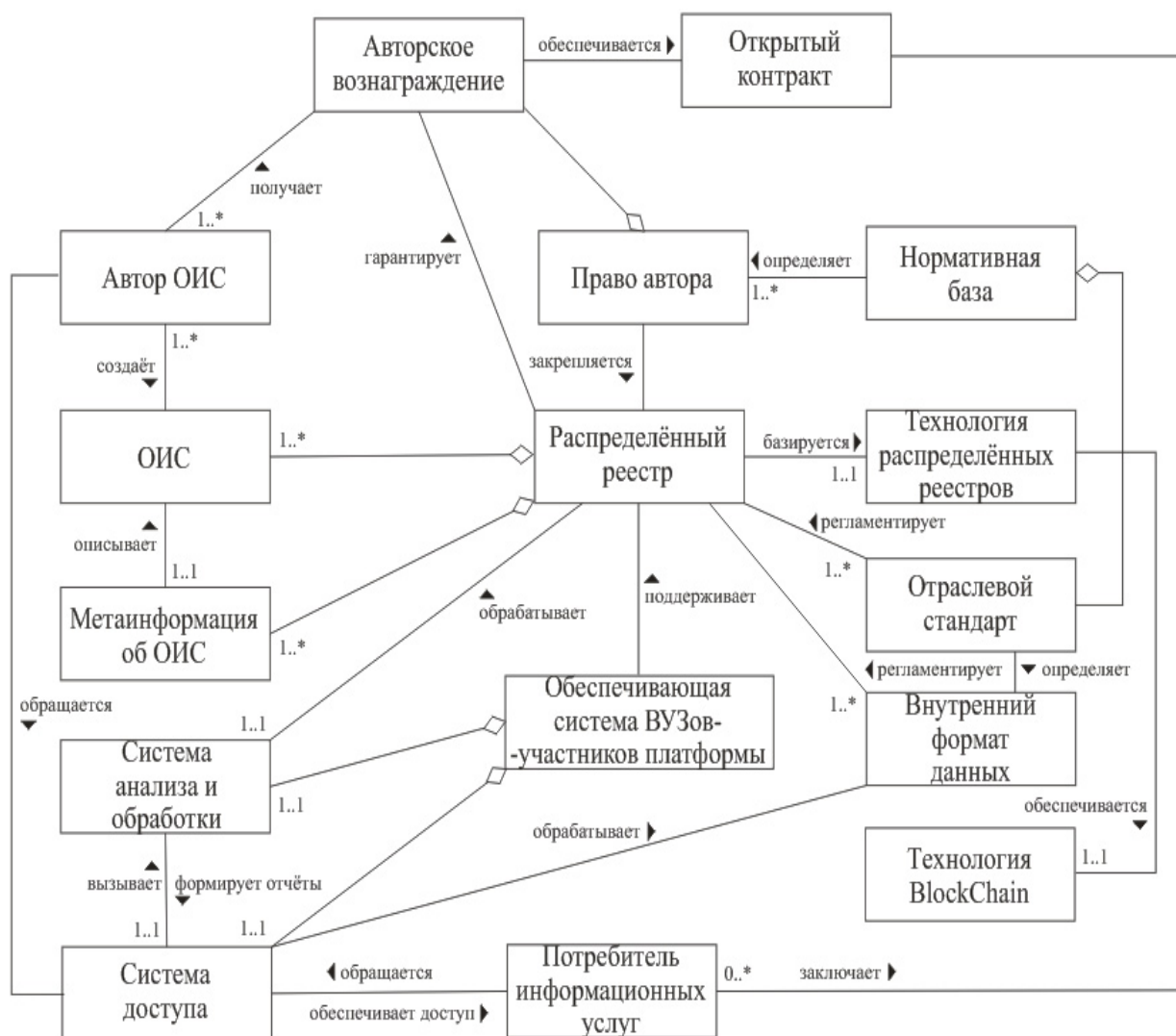


Рис. 1. Концептуальная модель предметной области [4]

⁹ IPC (IPUniversity IPO Code) – уникальное цифробуквенное обозначение (код), присваиваемое каждому ОИС (IPO, Intellectual Property Object) после завершения процесса его депонирования в информационной системе IPUniversity, которое является публичным свойством ОИС в реестре и может использоваться в качестве его краткого обозначения / ссылки.

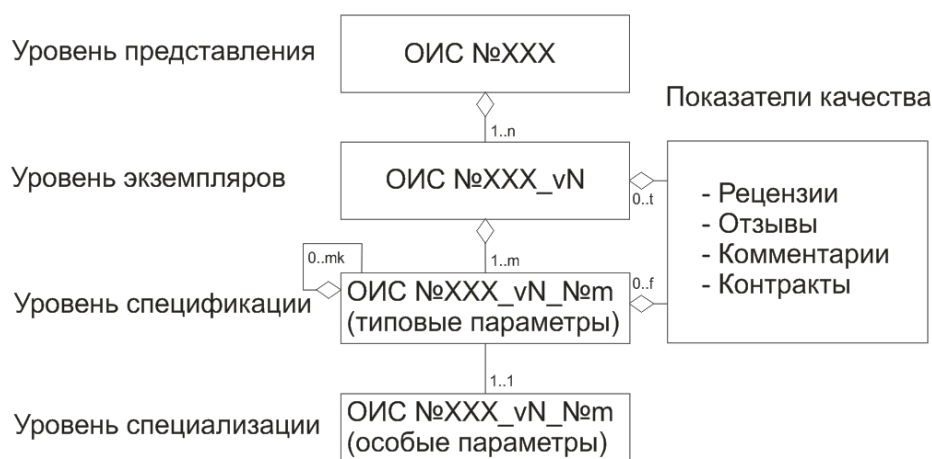


Рис. 2. Укрупненная модель агрегации-декомпозиции структуры метаданных ФЭР

Уровень спецификации включает описание внутренней гетерогенности (фрагментов ОИС) и связей с внешними сущностями (показателей качества). Включает следующие метаданные: идентификатор уровня экземпляров, доля вклада каждого из авторов, собственное заглавие (опционально), тип лицензии распространения фрагмента (опционально), вид и формат хранимой информации, множество ключевых слов (опционально), порядковый номер фрагмента в структуре экземпляра, а так же ссылки на структурные фрагменты элемента (каждый имеет аналогичную структуру).

Уровень специализации определяет дополнительные свойства, уникальные для данного подвида ФЭР, а так же специальные параметры.

Показатели качества – множество откликов пользовательского сообщества на содержимое ОИС, включая факты приобретения и цитирования, которые проводятся в реестре через смарт-контракты.

Отметим, что на уровне спецификации задаются данные (слои) как о структуре ФЭР, так и о правовом статусе (распределение вклада соавторов в ОИС и указание заимствований). Структурный слой может быть задан в виде совокупности описаний фрагментов объектов и ссылок как на такие же описания фрагментов, так и на файлы-ресурсы (собственно ОИС). Рассмотрим типы организации данных о структуре ФЭР подробнее.

Одноуровневый простой – цельные файлы ОИС, имеющие описание в метаданных фрагментов на уровне авторского слоя и/или слоя заимствований (структурный фрагмент один). Например, блок-схема алгоритма работы программы для ЭВМ, файл с описанием фрагмента ДНК или модель-программа работы станка с ЧПУ для изготовления какого-либо объекта (G-code).

Одноуровневый составной – цельные файлы ОИС, имеющие описание в метаданных фрагментов на уровне структуры, авторского слоя и/или слоя заимствований (структурные фрагменты ссылаются на части ресурсного файла). Например, препринт научной статьи, представленной в виде одного файла-макета.

Одноуровневый детализованный – цельные файлы ОИС, имеющие описание в метаданных фрагментов на уровне структуры, авторского слоя и/или слоя

заимствований, но дополненные ссылками отдельных фрагментов на собственные ресурсные файлы, не имеющие самостоятельного ИРС-номера (на изображения в высоком разрешении, полные протоколы экспериментальных первичных данных и т.п.). Например, препринт монографии, дополненный таблицами исходных данных, по которым стоились отдельные графики и диаграммы.

Двухуровневый простой – множество файлов (коллекция объектов или их архив), составляющих единый ОИС с общим описанием структурного и правового слоев в метаданных (общий ИРС-номер для всего множества). Например, обучающая выборка файлов-изображений для системы технического зрения или атлас (набор изображений-карт) определенной территории.

Двухуровневый составной – множество файлов, составляющих единый ОИС, данные о правовом слое в котором имеют индивидуальное описание для каждого файла (общий ИРС-номер для всего множества). Например, библиотека объектов проектирования.

Двухуровневый детализированный – множество файлов, составляющих ОИС, каждый фрагмент которого ссылается на внешний ОИС с собственным ИРС-номером в реестре. Например, препринт сборника научных статей, каждый фрагмент которой был ранее депонирован авторами независимо.

Многоуровневый – множество файлов, описанных в виде иерархии вложенных друг в друга ИС (произвольное объединение типов через механизм ссылочности). Например, электронный образовательный ресурс, включающий лекции, практические задания, примеры, контрольно-измерительные материалы и пр.

Для отдельных типов ФЭР (преимущественно текстовой природы) предусмотрена возможность вносить дополнения в файл-ресурс с помощью механизма расстановки меток, конкретизируя структуру, авторский вклад и области заимствования.

Теперь остановимся на понятии *авторский вклад* – это абсолютный показатель, отражающий долю участия (вклад) автора в создании коллективного объекта интеллектуальной деятельности. Метаданные об авторском вкладе в элементы ФЭР формируют авторский слой описания и предполагают три стратегии заполнения:

По роли (ролевая) – авторский вклад распределяется между авторами условно, исходя из различной роли каждого при создании ОИС (без привязки к фрагментам). Например, при секвенировании последовательности нуклеотидов указывается доля, отводимая руководителю проекта, оператору секвенатора, прочим участникам коллектива, непосредственно задействованных в создании файла ОИС.

По общему вкладу (долевая) – авторский вклад распределяется на весь ОИС, имеющим однородное ролевое исполнение (общее для всех фрагментов). Например, авторы статьи договариваются, что при депонировании распределят между собой авторство в равной доле.

По индивидуальному вкладу (индивидуализированная) – авторский вклад специфицируется в метаданных для каждого фрагмента отдельно. Например, главы в учебнике или коллективной монографии распределены между авторами в привязке к каждому фрагменту ОИС.

Для завершения формализации ФЭР как ОИС, для которого в последствии должен рассчитываться индекс цитируемости, следует дополнительно указать слой заимствований. Он задаётся массивом векторов, включающих: порядковый номер источника библиографии в рамках статьи; порядковый номер вхождения цитаты в текстовый ФЭР; ссылка на электронный первоисточник; тип цитирования; метод формирования. Таким образом, одни и те же метки о заимствовании могут встречаться в документе многократно, но с индивидуальными указаниями на тип цитирования. Допустим случай, когда тип цитирования может быть автоматически откорректирован сервисом проверки на наличие некорректных заимствований в случае выявления нарушения авторских прав.

Представленная в данной статье модель организации фрагментированных электронных ресурсов имеет в составе реестра ОИС следующий перечень актуальных *аналитических задач*:

- оценка авторской чистоты;
- оценка индекса цитируемости ФЭР;
- оценка индекса цитируемости фрагмента;
- оценка рейтинга автора по портфелю ОИС, включающего ФЭР;
- построение дерева связей ОИС с другими объектами реестра;
- построение карты портфеля ОИС автора или группы авторов;
- прогнозирование и выявление тенденций по направлению работы автора или их группы;
- выявление «каруселей цитирования».

На данном этапе реализована техническая часть реестра, находящегося по адресу ipuniversity.ru, которая позволяет депонировать целостные и фрагментированные электронные ресурсы. Методически проработаны и реализуются указанные аналитические выше функции, относящиеся к модулю аналитики системы [4]. Параллельно с этим идёт работа над созданием группы проектов Госстандартов серии 7.0X «Фрагментированные электронные ресурсы», ориентированной на применения в научной и университетской среде при депонировании объектов интеллектуальной собственности.

Работа выполнена в рамках государственного контракта № 14.596.11.0042.

Список литературы

1. Гражданский кодекс Российской Федерации. Ч. 4 (18.12.2006, №230 ФЗ).
2. Конкурсная документация лота 2017-14-596-0008 «Создание и запуск цифровой платформы обмена знаниями и управления авторскими правами» [Электронный ресурс]: режим доступа – http://fcpir.ru/participation_in_program/contests/list_of_contests/1_published/2017-14-596-0008.
3. ГОСТ Р 57100-2016 «Системная и программная инженерия. Описание архитектуры». – М.: Стандартинформ, 2016. – 36 с.
4. Углев В.А., Барышев Р.А. Сервисы интеллектуального анализа данных в проекте цифровой платформы обмена знаниями и управления авторскими правами // Робототехника и искусственный интеллект: материалы IX Всероссийской научно-технической конференции с международным участием. – Красноярск: ЛИТЕРА-принт, 2017. – С. 141-146.

ОЦЕНКА ПОКАЗАТЕЛЯ ДОЛИ АВТОРСКОГО ВКЛАДА, АВТОРСКОЙ «ЧИСТОТЫ» И ЦИТИРУЕМОСТИ ФРАГМЕНТИРОВАННЫХ РЕСУРСОВ

В.А. Углев, Я.А. Феодоров

Сибирский федеральный университет, uglev-v@yandex.ru, feodorov.yan@mail.ru

Цитируемость – это один из главных показателей в наукометрии и библиометрии, позволяющий количественно оценить вклад и востребованность отдельной научной работы в рамках определенного научного направления, опираясь на анализ статистики о заимствованиях. Научные работы (статьи, монографии, книги и т.п.) могут иметь одного и более соавторов, являясь *объектами интеллектуальной собственности* (ОИС), которые овеществляют результат научной деятельности. На базе показателя цитируемости множества работ автора (*портфеля* ОИС) можно осуществить расчёт рейтинга автора в различных контекстах (отраслевом, институциональном, проектном, тематическом, географическом, временном и пр.), а так же получить оценку производных показателей научных изданий (так называемые импакт-факторы). На сегодняшний день существует много различных методик расчёта импакт-факторов, которые базируются как исключительно на показателях цитируемости [1], так и на комплексном подходе (например, см. [2]). Таким образом, исходными единицами анализа становятся сами публикации и их связи с цитируемыми источниками, оформленными, как правило, в виде списка литературы. Пока ОИС рассматривается в виде неделимого объекта, а вклад авторов равноценен, расчёт показателей цитируемости не вызывает затруднений. Но эти допущения как правило не отражают реальные обстоятельства создания ОИС и приводят к существенной погрешности как в оценке заслуг соавторов¹⁰, так и в оценке уровня цитируемости¹¹. Рассмотрим задачу *оценки показателя цитируемости* для таких ОИС, которые имеют фрагментацию по структуре, авторскому вкладу и характеру цитирования внешних источников.

Определим понятие *фрагментированный электронный ресурс* (ЭФР) – это электронный информационный ресурс, который имеет неоднородную (фрагментированную) структуру, определяемую, как правило, семантикой объекта, типом включенных данных и/или правовым статусом фрагментов¹². Упуская не существенные для темы данной статьи параметры ФЭР, хранящиеся в базе (реестре) ОИС в виде совокупности метаданных, выделим те, которые позволяют обеспечить дальнейшее определение показателя цитируемости. Это данные, описывающие такие сущности, как соавторы, фрагменты ОИС, распределение долей вклада соавторов

¹⁰ Например, когда в соавторы добавляются должностные лица, принимающие формальное участие в создании конкретных ОИС, которые получают значительный совокупный рейтинг в рамках организации (в отличие от реальных авторов работ).

¹¹ Например, когда обзорная статья, базирующаяся на конкретных результатах других работ, «присваивает» себе факты ссылок на эти результаты, а заимствованные фрагменты первоисточников (особенно если это графики, схемы, таблицы, формулы, формулировки теорем и п.т.) эти ссылки не фиксируют.

¹² Здесь и далее будут рассматриваться электронные ресурсы, но, по сути, они могут иметь и не электронную природу (например, патент или книга), но быть зарегистрированной в соответствующей информационной системе (реестре). Кроме того, ОИС и ФЭР будут использоваться как синонимы.

(пофрагментно), цитируемые первоисточники и области заимствования. Для идентификации самого ОИС задаётся заглавие (обязательно русский и английский), идентификатор (например, адрес URL, коды DOI и/или IPC¹³). Для соавторов задается последовательность (в соответствии с библиографической записью). Для первоисточников – ссылки на адреса оригинальных текстов, их библиографические записи или коды. Остальные показатели распишем подробнее.

Фрагменты ОИС (структурный слой в метаданных) описываются массивом значений, определяющих: идентификатор фрагмента (порядковый номер), признак делимости фрагмента (логическое значение), вид фрагмента (1 – служебное метаописание, 2 – информационный блок, 3 – блок библиографии), наличие у фрагмента отдельной спецификации в метаданных данного ОИС или внешней сущности (логическое значение), а так же доля фрагмента в общем объеме работы (исключительно среди фрагментов вида 2).

Данные о цитировании (слой заимствований в метаданных) описываются массивом значений, определяющих: порядковый номер источника из библиографии, порядковый номер цитаты источника (минимум один раз), *тип цитирования* (1 – прямое цитирование, 2 – прямое цитирование с дополнением, 3 – не прямое цитирование), принадлежность цитаты фрагменту ФЭР, процент занимаемого цитатой объема внутри фрагмента. При таком кодировании будем полагать, что цитировании фрагмента первого типа даст коллективу соавторов 0% авторства на цитируемый фрагмент, второго – 50%, а третьего – 90%.

Распределение долей вклада соавторов (авторский слой в метаданных) описывается массивом значений (пофрагментно), определяющих: идентификатор фрагмента (порядковый номер), доля вклада процитированных заимствований во фрагменте (вычисляется автоматически, как % символов или внедренных объектов), доли вклада каждого соавтора (%) в заданной последовательности. При таком кодировании будем полагать, что суммарный вклад всех авторов во фрагмент вида 1 и 3 всегда принимает значение 0. При этом суммарный вклад всех авторов во фрагмент вида 2 принимает значение 100 при любых условиях на момент ввода данных о депонируемом текстовом ФЭР (далее пропорционально уменьшается на долю цитирования).

Рассмотрим пример описания всех трех слоев для [3], исходя из того, что статья имеет двух соавторов, 8 фрагментов, 8 источников библиографии (одноуровневый составной ФЭР с индивидуализированной стратегией заполнения данных по авторскому слою). Закодированные данные представлены в табл. 1, где пятые позиции векторов данных структурного слоя и слоя заимствований (проценты от объема), а так же второй столбец авторского (доля) можно автоматически рассчитать на основании анализа файла ОИС и имеющегося метаописания.

Таблица 1

Данные из метаописания ОИС для [3]

¹³ Включая вариант ссылки на фрагмент ФЭР через код IPC с соответствующим префиксом.

Структурный слой	Слой заимствований	Авторский слой
1; 0; 1; 0; 0	1; 1; 3; 2; 3	1; 0, 0, 0
2; 0; 2; 0; 21	2; 1; 3; 2; 3	2; 1.3, 88.83, 9.87
3; 0; 2; 0; 19	3; 1; 3; 2; 5	3; 0, 100, 0
4; 0; 2; 0; 34	4; 1; 3; 2; 2	4; 0.8, 89.28, 9.92
5; 0; 2; 0; 16	5; 1; 3; 4; 2	5; 0, 100, 0
6; 0; 2; 0; 10	6; 1; 3; 4; 3	6; 15, 68, 17
7; 0; 1; 0; 0	7; 1; 3; 4; 3	7; 0, 0, 0
8; 0; 3; 0; 0	8; 1; 2; 6; 30	8; 0, 0, 0

Для дальнейшего анализа авторского вклада и цитируемости потребуются ввести ряд расчетных показателей. Первым из них будет вклад заимствования (цитаты) j во фрагменте i ¹⁴, рассчитываемый по (1).

$$c_{i,j} = v_{i,j} * k, \quad (1)$$

где $v_{i,j}$ – пятое значение в векторах слоя заимствований, k – весовой коэффициент, определяемый типом цитаты (третье значение в векторах). Согласно принятым соотношениям, для первого типа цитирования значение k будет равно 1, для второго – 0.5, а для третьего – 0.1. Тогда вклад всех цитат в первом фрагменте нашего примера будет равен $c_1 = 0*1 + 0*0.5 + (3+3+5+2)*0.1 = 1.3\%$. Следовательно, для оценки вклада соавтора с порядковым номером m во фрагмент i относительно общего объёма работы g_i (пятое значение векторов структурного слоя) будет использована формула (2).

$$d_{i,m} = g_i * a_{i,m} / 100, \quad (2)$$

где $a_{i,m}$ – вклад автора (одно из значений авторского слоя с позиции 3, согласно номеру соавтора). Для рассматриваемого примера, вклад первого и второго соавторов в первый информационный фрагмент будет равен $d_{1,1} = 21*88.83/100 = 18.65\%$ и $d_{1,2} = 21*9.87/100 = 2.07\%$ соответственно.

На основании (1) и (2) можно рассчитать общий *объём заимствований*, как сумму c_i , и доли *авторского вклада*, как сумму $d_{i,m}$. В примере показатель заимствования равен 2.045%, а авторский вклад – 90.8095% и 7.1455% для первого и второго соавтора соответственно. В идеале, их сумма должна составлять 100%. Обратный показатель объёму заимствований является *объём оригинальности*.

Практика проверки корректности заимствований показывает, что соблюдение этики цитирования далеко не всегда соблюдается даже в научной среде. А это на прямую влияет на точность расчёта оценки показателя цитируемости. Поэтому введем

¹⁴ Нумерация фрагментов при расчётах ведётся исключительно для блоков информационного типа (закодированных как 2). В примере это фрагменты со 2-го по 6-й, имеющие номера с 1-го по 5-й соответственно. Следовательно, все формулы применимы только для блоков данного типа с соответствующими значениями i .

показатель авторской «чистоты» Ω , позволяющий дать количественную оценку совокупной доли оригинальных авторских результатов и цитирований к общему объёму ОИС, включающему данные о выявленных недобросовестных заимствованиях (по аналогии с термином «патентная чистота»):

$$\Omega = \sum_i \sum_m (d_{m,i} + c_i) - \delta, \quad (3)$$

где δ – интегральный показатель, количественно оценивающий объём некорректных цитат и факты плагиата¹⁵.

Теперь перейдём непосредственно к рассмотрению показателя цитируемости ФЭР. Пусть имеется ссылка сторонних авторов на статью [3] (а точнее, с прямым указанием на на фрагмент 3¹⁶, представленный в виде рисунка-схемы). Покажем, какой вклад сделает эта ссылка при условии, что исходная статья была описана различными способами и что цитирования оформлены корректно (см. табл. 2).

Таблица 2

Приращение вклада показателя цитирования фрагмента из [3] для его авторов

Тип цитирования	Фрагментация отсутствует (традиционный подход)	Стратегия фрагментации на авторском слое		
		Ролевая	Долевая	Индивидуализированная
		Общее соотношение долей авторов по статье 91 / 7.1		Согласно табл. 1
Прямое (схема продублирована)	1 / 1	0.91 / 0.071	0.91 / 0.071	1 / 0
Прямое с дополнением (схема доработана)	1 / 1	0.91 / 0.071	0.91 / 0.071	1 / 0
Не прямое (схема упомянута)	1 / 1	0.91 / 0.071	0.91 / 0.071	1 / 0

Как видно из табл. 2, тип цитирования никак не влияет на показатель цитируемости авторов контента. Очевидно, что если в новой статье ссылку сделает один из соавторов, то она отразится только в статистике самоцитирования, никак не изменив общий показатель цитируемости.

Теперь покажем, какой вклад сделает эта ссылка при условии, что цитируемый фрагмент (изображение-схема из [3]) был позаимствован из второй статьи ещё одним (третьим) коллективом авторов, а затем четвертым коллективом (последовательное цитирование, согласно табл. 3), при условии, что все четыре работы были описаны как ФЭР с индивидуализированным авторским слоем.

¹⁵ При автоматизированной оценке ОИС на наличие плагиата можно использовать данные сервиса <https://www.antiplagiat.ru/>, протокол которой можно детально проанализировать, например, с помощью экспертной системы. Очевидно, что данная возможность доступна лишь для ЭФР текстовой природы. Для прочих видов информации создание технологии на проверку на наличие некорректных заимствований является перспективной областью исследований.

¹⁶ Для указания фрагмента используется префикс в коде ИРС при оформлении ссылки.

Таблица 3

Приращение вклада к показателю цитирования фрагмента из [3] для авторов

Тип цитирования	Авторы фрагмента (согласно табл. 1)	Авторы, заимствующие из оригинала (на всех)	Авторы, заимствующие из вторичного источника (на всех)
	Исходная статья	Статья 2	Статья 3
Прямое (схема продублирована)	3 / 0	0	0
Прямое с дополнением (схема доработана)	3 / 0	2	1
Не прямое (схема упомянута)	1 / 0	1	1

Из таблицы следует, что простое копирование контента при корректном описании цитат даёт рост показателя цитируемости только первоисточнику. Но наивно полагать, что без автоматического механизма проверки ОИС на авторскую «чистоту» и публичного «штрафования» за плагиат цитирующие авторы будут заинтересованы в данном подходе.

Рассмотренный в статье подход к оценке показателя цитируемости ориентирован на применение в информационных системах, оперирующих научными электронными ресурсами, допускающими фрагментированное описание. Как правило, сам показатель цитируемости редко применим для непосредственной оценки активности учёного и претерпевает дальнейшее преобразование (например, в процессе расчёта индекса Хирша) или включается в состав более сложных показателей [2]. Кроме того, данные о цитируемости фрагментов научных работ и статистике обращений к ним интересны для анализа в личном профиле учёного [4], порождая задачи прогнозирования, когнитивной визуализации, определения рейтингов и поиска по портфелю объектов интеллектуальной собственности.

Список литературы

5. Писляков В.В. Наукометрические методы и практики, рекомендуемые к применению в работе с российским индексом научного цитирования [Электронный ресурс]. – Режим доступа: <http://www.elibrary.ru/projects/citation/docs/scientometrics.pdf>.

6. Редькина Н.С. Модель многоуровневого изучения результативности научных исследований [Электронный ресурс]. – Режим доступа: <http://www.ict.nsc.ru/ws/YM2006/10617/Redkina.pdf>.

7. Углев В.А., Барышев Р.А. Сервисы интеллектуального анализа данных в проекте цифровой платформы обмена знаниями и управления авторскими правами // Робототехника и искусственный интеллект: материалы IX Всероссийской научно-технической конференции с международным участием. – Красноярск: ЛИТЕРА-принт, 2017. – С. 141-146.

8. Углев В.А., Егорова Д.В. Анализ публикационной активности учёного и методы Graph Mining // Робототехника и искусственный интеллект: Материалы VIII Всероссийской научно-технической конференции с международным участием. – Красноярск: Сиб. федер. ун-т., 2016. – С. 162-168.

ПРИМЕНЕНИЕ КОМБИНАЦИИ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ И КЛАССИФИКАЦИИ ЛЮДЕЙ НА ВИДЕОИЗОБРАЖЕНИИ

И.С.Фомин¹, С.Р.Орлова¹, А.В.Бахшиев^{1,2}

¹ Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК), i.fomin@rtc.ru,

² Санкт-Петербургский политехнический университет Петра Великого (СПбПУ),
bakhshiev_av@spbstu.ru

На сегодняшний день видеонаблюдение при помощи камер является стандартом при обеспечении безопасности территории большинства предприятий от незаконного проникновения и иных противоправных действий, таких как кража или вандализм. Камеры устанавливаются также в помещениях предприятий, где позволяют наблюдать за действиями сотрудников или опасными областями. Чаще всего данные с камер собираются на пульт, где просматриваются оператором, все вместе, или только некоторые. Из-за большого числа камер и долгих периодов, в которые ничего не происходит, оператор может пропустить событие или среагировать с задержкой. Поэтому получили развития системы видеонаблюдения различной сложности. Они обрабатывают данные с каждой камеры постоянно и не могут утратить концентрацию, поэтому повышают шансы обнаружить нежелательное событие. Самые простые системы позволяют определить наличие движения в кадре (такие алгоритмы сейчас интегрированы в большинство камер), более сложные позволяют идентифицировать и сопровождать отдельные объекты. Важной частью системы видеонаблюдения является алгоритм обнаружения объектов.

Для обнаружения и сопровождения объектов используются разные методы. Многие системы видеонаблюдения используют подход с вычитанием фона [1], а для обнаружения объектов, вид которых известен могут быть использованы различные детекторы, основанные на методах машинного обучения, например, каскадный детектор [2]. В последнее время активно развиваются сверточные нейронные сети, которые показали хорошие результаты в классификации объектов, а позднее были распространены и на обнаружение.

В данной работе рассматривается использование комбинации нейросетевого алгоритма обнаружения объектов на изображении и сверточной нейронной сети для классификации объектов. Первый алгоритм – нейронная сеть для обнаружения объектов определенного класса (людей в нашем случае), которая, тем не менее, неспособна делать это без ошибок. Достаточно часто встречаются ошибки первого и второго рода, кроме того, показатель уверенности обнаружения зачастую выше у объектов, которые обнаружены ошибочно, поэтому результаты трудно отфильтровать. Для исправления этого недостатка предлагается использовать вторую сеть, которая будет обучена на результатах работы детектора и позволит отделить правильные обнаружения от неправильных. Предлагается использовать результаты обнаружения объектов с помощью детектора, чтобы сформировать набор данных для обучения классификатора, разметив верные и ложные обнаружения, это быстрее, чем рисовать рамку вокруг

каждого объекта вручную. Используя эти данные обучается классификатор, способный успешно разделять правильные результаты детектора от ложных. В результате получается алгоритм для обнаружения объектов, дающий намного более точный выход без задействования дорогих технических средств для обучения детектора.

Еще одним решением является доучивание детектора на интересующих объектах и сложных отрицательных примерах, но это сопряжено с трудностями, так как требует большой размеченной базы с положительными и сложными отрицательными примерами из материала, на котором сети предстоит работать, но создание такой разметки длительный и трудоемкий процесс. Кроме того, существующие детекторы требуют для обучения дорогостоящее оборудование – графические процессоры с большим количеством видеопамяти. Предложенный классификатор может быть обучен на том же оборудовании, на котором работает детектор.

Рассмотрим более подробно существующие алгоритмы обнаружения объектов, основанные на нейронных сетях. Один из первых достаточно эффективных детекторов объектов, использующих нейронные сети, носит название R-CNN [3] и использует нейронную сеть как средство классификации объектов. Он использует два различных алгоритма для выделения зон интереса на изображении, а по результатам работы этих алгоритмов выполняет обработку нейронной сетью, которая классифицирует области изображения, попавшие в зоны интереса. Объединение этих двух методов и нейронной сети для классификации и представляет собой R-CNN. Позже появилось еще несколько итераций данного метода, где он был доработан следующим образом. R-CNN выполняет независимую обработку каждого прямоугольника, выделенного внешним методом, а считать все слои нейронной сети каждый раз заново очень неэффективно. Поэтому в новой модификации Fast R-CNN [4] сначала нейронная сеть вычисляет все сверточные слои для всего изображения, а потом полученные области интереса проецируются на последний сверточный слой и поступают на вход полносвязной части – классификатора. Последнее улучшение – замена внешнего средства для выделения возможных положений объектов специально созданной нейронной сетью, которая разделяет веса с основной классифицирующей частью и учится вместе с ней в процессе предъявления примеров, Faster R-CNN [5]. Это повысило точность получаемых ограничивающих прямоугольников, снизило их количество и количество расчетов классификации, а также увеличило производительность, так как нейронная сеть рассчитывается на графическом ускорителе заметно быстрее, чем использованные ранее методы на центральном процессоре.

Сеть YOLO (You Only Look Once) [6] не использует принцип скользящего окна в виде, в котором он используется в большинстве детекторов (перемещение окна с небольшим шагом и каскадная классификация). Вместо этого изображение разбивается на сетку с некоторым заранее определенным числом ячеек по горизонтали и вертикали. Перед началом работы задается набор якорных прямоугольников для каждой ячейки, после чего для каждого из прямоугольников выполняется проецирование на выход последнего сверточного слоя и классификация полносвязным слоем. Набор параметров для выбора якорных прямоугольников может быть выбран вручную или вычислен при

помощи специального средства исходя из ограничивающих прямоугольников, которые встречаются в обучающем наборе. Алгоритм не пытается выполнять дополнительное уточнение положения объекта в специальных слоях, только проверяет, содержит ли тот или иной прямоугольник объект или фон. Такой подход не снижает точности и полноты обнаружения.

Кроме того, позже появилась версия YOLOv2 [7], готовая реализация которой была использована в работе. В ней авторами исходной статьи были выполнены несколько улучшений и дополнений относительно первой версии. Во-первых, вторая версия использует нормализацию наборов обучающих изображений, поступающих на вход сети на каждой итерации обучения (нормализация батчей или *batch normalization*). Это позволило увеличить показатель mAP (*mean average precision*, метрика качества обнаружения объектов из PASCAL VOC CHALLENGE, общепринятый показатель качества) на 2%. Увеличение разрешения классификатора с классических 256x256 (разрешение AlexNet, которое впоследствии стало очень популярным при разработке классификаторов) до 448x448 позволило повысить качество еще на 4%. Использование набора якорных прямоугольников, полученных специальным средством, сократило время определения входных данных для классификатора и позволило системе вместо этого проверять более 1000 прямоугольников на каждом изображении. Этот факт значительно повысил полноту обнаружений объектов. Пересчет координат прямоугольника из координат относительно левого верхнего угла изображения в координаты относительно центра рамки упростил параметризацию расчета и повысил инвариантность сети к положению объекта на изображении, что также положительно сказалось на показателе качества обнаружения. Карта признаков с последнего слоя изображения была расширена специально преобразованными признаками с предыдущего слоя. Это позволило обнаруживать меньшие по размеру объекты, чем ранее. Также применено случайное изменение размера входного изображения в процессе обучения, поэтому сеть стала способна качественно обнаруживать объекты на изображениях разного размера, что повысило ее применимость. Изменение входного размера позволяет пользователю подобрать баланс между производительностью сети и выходным показателем качества.

В качестве детектора выбрана нейронная сеть YOLOv2, описанная выше. Данная сеть имеет несколько возможных реализаций под разные фреймворки для организации нейронных сетей. Есть версии под darknet, Caffe, Caffe2, TensorFlow&Keras, PyTorch и другие. В качестве реализации был выбран вариант, основанный на TensorFlow&Keras. Он находится в свободном доступе и выложен на GitHub [8]. Кроме того, что имеется готовая реализация, есть также готовые наборы весов нейронной сети, позволяющие обнаруживать различные классы объектов, в том числе и людей. При обнаружении людей очень сложно определить обнаружен ли человек, имея информацию только о положении объекта и уверенности сети. В сети отсутствуют какие-либо средства сопровождения объектов и нет возможности исходя из того, что объект был ранее идентифицирован, как человек, сделать какие-либо выводы о том, что объект по-прежнему относится к классу «человек». Кроме того, уверенность сети сильно колеблется и некоторые объекты могут быть по ошибке идентифицированы как человек

с большей уверенностью, чем настоящий человек. Для того, чтобы повысить стабильность сопровождения людей на видео, предприняты два улучшения. Во-первых, добавлена небольшая нейронная сеть, которая выступает в качестве бинарного классификатора, что позволяет успешно отсеять большинство ложных обнаружений. Во-вторых, добавлен алгоритм фильтрации, позволяющий отсеять единичные ошибки сети. Алгоритм предполагает, что камера фиксирована, и человек мало смещается между двумя соседними кадрами. Вместе эти два улучшения позволяют отсеять большинство ложных обнаружений, не потеряв существенно в показателях полноты и точности обнаружения.

Сеть классификатора представляет собой аналог классических нейронных сетей для классификации объектов. Схема сети представлена на рисунке ниже. Следует сказать, что размер исходного изображения в 96 пикселей выбран абсолютно произвольно, с учетом того, что сеть должна работать быстро, требовать мало памяти видеокарты и процессорного времени для расчета. Также входное изображение должно быть достаточно большим чтобы отражать внешний вид объекта в достаточной для нейронной сети степени. Параметры слоев выбраны исходя из показателя точности, на момент написания статьи на тестовом наборе данных она составляла 95,35%, этого достаточно для правильной идентификации большинства сложных для детектора примеров.



Рис. Схема сети бинарной классификации. Пунктирными линиями показан переход между слоями, в которых при обучении отбрасывается часть нейронов (dropout). Форма сверточных слоев записана как (число ядер)x(ширина ядра)x(высота ядра).

Фильтрация выполняется с учетом того, что камера фиксирована, а человек между двумя кадрами на большое расстояние переместиться не может. Если в кадре идентифицирован некий объект, то он сначала подвергается обработке бинарным классификатором, после чего выполняется последовательный поиск ближайшего обнаруженного объекта на предыдущих кадрах. Если объект обнаруживается нестабильно, то есть нет хотя бы 5 последовательных кадров с обнаружениями, то он относится к ошибкам. Если несколько последовательных обнаружений есть, то вычисляется средняя уверенность детектора (уверенность того, что обнаруженный объект есть человек) и процент кадров на которых бинарный классификатор отнес объект к классу человека. Если средняя уверенность или процент верных обнаружений превышают соответствующие пороги, то объект определяется как человек.

В качестве обучающих данных были использованы результаты работы детектора людей на нескольких видеозаписях с фиксированной камеры, на которых человек перемещается в интерьере цеха. Для каждого из результирующих прямоугольников было указано, имеется в прямоугольнике человек или нет. После этого размеченные фрагменты изображений были дополнены случайно выбранными отрицательными примерами и использованы для обучения показанной выше нейронной сети. Набор был разделен на обучающий и тестовый. Всего в тестовом наборе 62766 изображений, из них 22680 положительных и 40086 отрицательных, в обучающем 131652 изображения, из них 49608 положительных и 82044 отрицательных. Количество примеров в обоих наборах данных указано с учетом аугментации, включающей в себя смещение прямоугольника относительно изначального положения и отражение изображений.

Результаты работы системы обнаружения объектов представлены в таблице. Значения, указанные в качестве результатов, рассчитываются исходя из стандартных значений для любого детектора и классификатора – количества верно и неверно обнаруженных позитивных и негативных примеров. Формулы расчета могут быть найдены, например, в [9]. Здесь точность это Precision (PPV, Positive Prediction Value), полнота это Recall (TPR, True Positive Rate), специфичность это Specifity (TNR, True Negative Rate), и аккуратность это Accuracy (ACC).

Таблица

Результаты обнаружения до и после использования классификатора

№	Без классификатора		С классификатором			
	Точность	Полнота	Точность	Полнота	Специфичность	Аккуратность
1	34,51	92,11	82,97	92,92	77,03	85,71
2	41,42	89,15	93,99	91,89	95,82	94,04
3	36,13	78,75	93,93	87,61	96,82	93,51

Проверка обоих вариантов выполнялась на трех экспериментальных видеозаписях, которые не использовались при формировании обучающего набора данных. Из полученных результатов можно сделать вывод, что нейронная сеть хороша как детектор лишь в том смысле, что обнаруживает большой процент нужных объектов (людей) на изображении, но большая часть выделенных сетью объектов все же являются ложными обнаружениями. Поэтому требуется дообучение сети, либо дополнительная классификация, по второму пути мы пошли в данной работе. Как видно из полученных результатов, дополнительная классификация и фильтры позволяют отсеять большую часть ложных обнаружений. Это неизбежно приводит к появлению ошибочных классификаций человека как фона, но точность при этом снижается незначительно.

В заключение следует отметить, что в данной работе был рассмотрен подход, при котором выход сложного детектора на основе нейронной сети дополнительно

обрабатывается небольшой сетью-классификатором и фильтруется по уверенности детектора и стабильности ответа классификатора на объект. Это позволяет существенно повысить точность получившейся комбинации двух сетей, что было продемонстрировано в результатах работы. Данный подход может быть использован в условиях, когда по тем или иным причинам невозможно дообучение нейросетевого детектора, но требуется повысить качество обнаружения. Такая ситуация возможна, так как для работы обученного детектора и простого классификатора, а также для обучения простого классификатора требуется существенно меньше ресурсов, чем для обучения или дообучения детектора полностью.

В качестве дальнейшей работы предполагается изменение структуры и дальнейшая настройка параметров классификатора для повышения точности. Также будут проведены эксперименты с классическими архитектурами классификаторов, для сравнения и возможной замены. Кроме того, предполагается использовать собранный материал для того, чтобы дообучить детектор и сравнить качество дообученного детектора с качеством работы комбинации детектора и классификатора.

Список литературы

1. Бахшиев А. В., Михайлов В. В. Программный комплекс для разработки систем технического зрения с динамически изменяемой модульной архитектурой //Робототехника и техническая кибернетика. – 2016. – №. 4. – С. 38-42.
2. Lienhart R., Maydt J. An extended set of haar-like features for rapid object detection //Image Processing. 2002. Proceedings. 2002 International Conference on. – IEEE, 2002. – Т. 1. – С. I-I.
3. Girshick R. et al. Region-based convolutional networks for accurate object detection and segmentation //IEEE transactions on pattern analysis and machine intelligence. – 2016. – Т. 38. – №. 1. – С. 142-158.
4. Girshick R. Fast r-cnn //Proceedings of the IEEE international conference on computer vision. – 2015. – С. 1440-1448.
5. Ren S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks //Advances in neural information processing systems. – 2015. – С. 91-99.
6. Redmon J. et al. You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. – С. 779-788.
7. Redmon J., Farhadi A. YOLO9000: better, faster, stronger. E-print: <https://arxiv.org/abs/1612.08242>
8. YOLOv2 in Keras and applications // URL: <https://github.com/experiencor/keras-yolo2>
9. Precision and Recall – Wikipedia // URL: https://en.wikipedia.org/wiki/Precision_and_recall

АНАЛИЗ ИНИЦИАТИВНЫХ НАУЧНЫХ ПРОЕКТОВ ПО ОБЛАСТИ ЗНАНИЯ "ФИЗИКА, АСТРОНОМИЯ", ПОДДЕРЖАННЫХ РОССИЙСКИМ ФОНДОМ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Р.А. Чиженкова
Институт биофизики клетки РАН, chizhenkova@mail.ru

Введение

В середине XX века понимание значения фундаментальной науки для судеб цивилизации привело в развитых странах к поиску новых подходов к организации науки [3- 5]. Философский анализ состояния науки в современном обществе уже проводился в ряде наших работ [9-14]. В настоящее время возникла система специальных фондов для поддержки научных коллективов и отдельных ученых [4, 7]. Российский Фонд Фундаментальных Исследований (РФФИ) был создан в 1992 г. [1]. Материалы относительно 20-летнего юбилея РФФИ отражены в специальном выпуске журнала "Вестник РФФИ". Однако в открытых публикациях не был проведен анализ научных направлений проектов, поддержанных Фондом, что послужило причиной проведения наших библиометрических исследований [12, 13].

Среди видов конкурсов, проводимых РФФИ, основным, несомненно, является конкурс инициативных (т.е. исследовательских) проектов [12, 13]. Поддержанные инициативные проекты касались нескольких областей знания: 1. "Математика, информатика, механика"; 2. "Физика, астрономия"; 3. "Химия и науки о материалах"; 4. "Биология и медицинская наука"; 5. "Науки о Земле"; 6. "Науки о человеке и обществе"; 7. "Информационные технологии и вычислительные системы"; 8. "Фундаментальные основы инженерных наук" [13]. Среди инициативных проектов указанных направлений по общим количественным аспектам примерно в равной степени достоверно лидировали таковые по областям знания "Физика, астрономия" и "Биология и медицинская наука" (соответственно 21.10% и 20.61% от общего числа) [13].

Настоящие исследования посвящены библиометрическому анализу инициативных проектов по области знания "Физика, астрономия", поддержанных РФФИ в течение 20-летнего периода.

Материалы и методы

Представленные здесь сведения основываются на материалах, опубликованных в Информационных бюллетенях (ИБ) РФФИ, выходящих раз в год и освещающих итоги прошедшего конкурса, которые позволяют рассмотреть количественные данные его результатов по разным областям знания.

Для каждого года составлялась обширная таблица с материалами конкурса проектов по области знания "Физика, астрономия" относительно чисел выделенных грантов и поданных заявок. На основе этого вычислялась процентная вероятность поддержки проектов. Проводили объединение результатов по анализируемым рубрикам за весь временной период и вычисляли статистическую значимость различия величин, составляющих эти совокупности. Для статистического анализа использовали сравнение двух выборочных долей вариантов. Кроме того, применяли корреляционный анализ.

Результаты и обсуждение

В целом, число поддержанных РФФИ инициативных проектов за 20-летний период достигало существенной величины - 58667. Из них 12376 проектов имели отношение к области знания "Физика, астрономия", что составляло 21.10% от общего числа. Динамика этих чисел за исследуемый период представлена в табл.1.

Таблица 1

Динамика чисел поддержанных инициативных проектов по области знания «Физика и астрономия» в течение 20-летнего периода

Годы	Инициативные проекты "Физика, астрономия"			
	Физика, астрономия	% от суммы	Все области знания	% от областей знания
1993	757	6.12	2845	26.61
1994	654	5.28	2856	22.90
1995	750	3.06	2674	28.05
1996	1290	10.42	4742	27.20
1997	559	4.52	2637	21.20
1998	656	5.30	2971	22.08
1999	593	4.79	3132	19.93
2000	639	5.16	2809	22.75
2001	649	5.24	3058	21.22
2002	632	5.11	3046	20.78
2003	598	4.83	2612	22.89
2004	586	4.73	2543	23.04
2005	650	5.25	2936	22.14
2006	523	4.23	3022	17.31
2007	513	4.15	2610	19.66
2008	506	4.09	2947	17.17
2009	491	3.97	2801	17.53
2010	448	3.62	2738	16.36
2011	462	3.73	2922	15.81
2012	420	3.39	2766	18.18
Сумма	12376	100	58667	21.10
Среднее	618.80	5.00	2933.35	21.10

Колебания годовых чисел грантов в основном захватывали диапазон от 420 до 757 при средней величине 618.80. При этом, как видно из таблицы 1, в 1996 г. наблюдался выраженный пик числа грантов - 1290, который составлял 10.42% от их общего числа за 20 лет. Следует отметить, что числа поддержанных проектов были выше в первое десятилетие. Тогда же отмечались и их наибольшие колебания, а также и их пиковое значение.

Числа поданных заявок по инициативным проектам в ИБ, к сожалению, приведены без учета данных в 1993-1995 и 2001 гг., т.е. в течение 16 из 20 лет. За указанный 16-летний период число поданных заявок по инициативным проектам составляло 150189. Из них 31096, т.е. 20.70% от общего числа, касались области знания "Физика, астрономия". Динамика чисел поданных заявок по области знания "Физика, астрономия" за исследуемый период отражена на в табл. 2.

Таблица 2

Динамика чисел поданных заявок по инициативным проектам
в области знания «Физика и астрономия»
в течение рассматриваемого периода

Годы	Инициативные проекты "Физика, астрономия"			
	Физика, астрономия	% от суммы	Все области знания	% от областей знания
1993				
1994				
1995				
1996	3824	12.30	15722	24.32
1997	2716	8.73	11129	24.41
1998	2406	7.74	9633	24.98
1999	2524	8.12	9968	25.32
2000	2031	6.53	8169	24.86
2001				
2002	1931	6.21	9056	21.32
2003	1766	5.68	8000	22.08
2004	1677	5.39	7839	21.39
2005	1860	5.98	8475	21.95
2006	1530	4.92	9149	16.72
2007	1487	4.78	7963	18.67
2008	1457	4.69	8530	17.08
2009	1527	4.91	8752	17.45
2010	1471	4.73	9066	16.23
2011	1521	4.89	9658	15.75
2012	1368	4.40	9080	15.07
Сумма	31096	100	150189	20.70
Среднее	1943.50	6.25	9386.813	20.70

Числа поданных заявок за один год колебались от 1368 до 2716 при среднем значении 1943.50. При этом в 1996 г. имел место их резкий пик - 3824, на который приходилось 12.30% от общего числа за 16 лет. Табл. 2 показывает, что в начале рассматриваемого периода числа поданных заявок были преимущественно выше, а также претерпевали наиболее выраженные колебания. Наименьшие числа поданных заявок отмечались в последние годы анализируемого периода.

Выявленные количественные характеристики выделенных грантов и представленных заявок позволяют определить вероятность (или возможность) поддержки посылаемых материалов. В течение 16 лет, по которым в ИБ даны числа заявок, из 31096 заявок по области знания "Физика, астрономия" было поддержано 9566 проектов, что составляло 30.76%. Динамика процентных чисел проектов по годам продемонстрирована на табл. 3.

Таблица 3

Динамика процентных чисел поддержанных инициативных проектов по области знания «Физика и астрономия» в течение рассматриваемого периода.

Годы	Инициативные проекты. Физика, астрономия.		
	Гранты	Заявки	% под- держанных
1993			
1994			
1995			
1996	1290	3824	33.73
1997	559	2716	20.58
1998	656	2406	27.27
1999	593	2524	23.49
2000	639	2031	31.46
2001			
2002	632	1931	32.73
2003	598	1766	33.86
2004	586	1677	34.94
2005	650	1860	34.95
2006	523	1530	34.18
2007	513	1487	34.50
2008	506	1457	34.73
2009	491	1527	32.15
2010	448	1471	30.46
2011	462	1521	30.37
2012	420	1368	30.70
Сумма	9566	31096	30.76
Среднее	597.875	1943.5	30.76

Процентные числа поддержанных проектов по области знания "Физика, астрономия" за рассматриваемые годы составляли от 20.58 до 34.95 при среднем значении 30.76%. Как следует из табл. 3, наибольшие колебания процентных чисел поддержанных проектов приходились на начальную часть 16-летнего периода. Наибольшие значения анализируемой величины отмечались в 1996 г. и в 2004-2008 гг., а наименьшие - в 1997 и 1999 гг.

Таким образом, за 20 лет своей деятельности РФФИ оказал значительную поддержку проведению исследований в области знания "Физика, астрономия", которой принадлежит существенная роль в развитии новейших технологий XXI-ого века [2, 6]. Проведено рассмотрение чисел выделенных грантов, чисел подаваемых заявок и процентных чисел поддержанных проектов, что позволило выявить следующие моменты.

Во-первых, число поддержанных инициативных проектов по области знания "Физика, астрономия" представляло весьма существенную величину - 12376, что составляло 21.10% от общего числа по всем восьми рассматриваемым областям знания. Доля поданных заявок проектов по области знания "Физика, астрономия" в общем числе таковых обладала близким показателем - 20.70%. Процентное число поддержанных проектов по области знания "Физика, астрономия" равнялось 30.76, что приближалось к аналогичному процентному числу в суммарных данных по всем областям знания - 31.45 [13].

Во-вторых, имели место выраженные колебания по годам чисел выделенных грантов, чисел подаваемых заявок и процентных чисел поддержанных проектов. Отмечалось некоторое сходство динамик чисел грантов и чисел заявок и

принципиальное их отличие от динамики процентных чисел поддержанных проектов. Однако во всех трех случаях колебания соответствующих величин отмечались преимущественно в начальной части анализируемого периода. При этом колебания чисел грантов были в 3.07 раз, сделанных заявок - 2.80 раз и процентов поддержанных проектов - в 1.69 раз.

Корреляционный анализ позволил выявить факт наличия некоторого параллелизма чисел выделенных грантов и чисел подаваемых заявок ($r = 0.87$; $p < 0.01$), что свидетельствует о значительной роли общего состояния внутренней возможности в поддержке научных проектов РФФИ. Кроме того, была установлена положительная корреляционная взаимосвязь чисел выделенных грантов, чисел подаваемых заявок и процентных чисел поддержанных проектов по области знания "Физика, астрономия" с соответствующими величинами суммарных данных по всем областям знания в конкурсе инициативных проектов ($r = 0.82, 0.87$ и 0.76 ; $p < 0.01$), что также позволяет предположить ведущее значение финансовых ресурсов РФФИ в результатах его деятельности.

Заключение

Организация РФФИ явилась одним из первых нововведений молодой РФ по опыту других развитых государств. За два десятилетия деятельности РФФИ была оказана существенная поддержка проведению фундаментальных исследований, в том числе и по области знания "Физика, астрономия". Для повышения возможности поддержки научных проектов весьма желательно увеличение финансового наполнения Фонда. На Общем собрании Российской академии наук в 2012 г., в своем выступлении В.В Путин дал высокую оценку работе Фонда и сказал о намерении правительства увеличить его ресурсное наполнение в ближайшие годы [8].

Список литературы

1. Алфимов М.В. Российский фонд фундаментальных исследований: десять лет служения российской науке // Вестник РФФИ. - 2002. - № 1(27). - С. 5-39.
2. Горохов В.Г., Сидоренко А.П. Роль фундаментальных исследований в развитии новейших технологий // Вопросы философии. 2009. - № 3. - С. 67-76.
3. Ефремов Ю.Н. Зачем нужна наука миру и России // Вестник РФФИ. - 2000. - № 1(19). - С. 40-43.
4. Коннов В.И. Самоуправление на "передовой": становление национального научного фонда США // Вестник РФФИ. - 2007. - №4(54). - С. 10-15.
5. Лебедев С.А. Праксиология науки // Вопросы философии. - 2012. - №4. - С. 52-63.
6. Липкин А.И. Две методологические революции в физике - ключ к пониманию оснований квантовой механики // Вопросы философии. 2010. - № 4. - С. 74-90.
7. Обама Б. Выступление в Национальной академии наук 27-ого апреля 2009 г. // В защиту науки. Бюл. № 6. Ред. Э.П. Кругляков. - М.: Наука, 2009. - С. 185-198.
8. Путин В.В. Выступление на Общем собрании Российской академии наук 22 мая 2012 // В защиту науки. Бюл. № 11. Ред. Э.П. Кругляков. - М.: Наука, 2012. - С. 7-14..
9. Чиженкова Р.А. Наука в современном мире // Актуальные проблемы социальной философии / ред. Э.В. Гирусов. - М., 1998. - С. 111-112.
10. Чиженкова Р.А. Библиометрический анализ нейрофизиологических аспектов действия неионизирующей радиации / Р.А. Чиженкова, А.А. Сафрошкина, Н.А. Слащева, В.Ю. Чернухин // Успехи современной биологии. - 2004. - Т. 124. - № 5. - С. 472-479.
11. Чиженкова Р.А. Динамика нейрофизиологических исследований действия неионизирующей радиации во второй половине XX-ого века. М.: Издат. дом Акад. Естествознания, 2012. - 88 с.
12. Чиженкова Р.А. Библиометрический анализ научных проектов, поддержанных Российским Фондом Фундаментальных Исследований за 20 лет: виды конкурсов // Успехи современного естествознания. - 2014. - №5. - С. 145-150.
13. Чиженкова Р.А. Российский фонд фундаментальных исследований. Библиометрический анализ поддержки научных проектов. Germany: Palmarium Academic Publishing, 2015. - 88 с.
14. Chizhenkova R.A. Bibliometrical review of neurophysiological investigation of action of non-ionized radiation in second half of the XXth century // Biophysics. - 2005. - Supplement. - № 1(50). - P. 163-172.

РАЗВИТИЕ ИММУННЫХ АЛГОРИТМОВ НА КЛАСС НЕЙРОСЕТЕВЫХ СИСТЕМ УПРАВЛЕНИЯ

О.И. Ширяева

Казахский национальный исследовательский технический университет им.К.Сатпаева,
oshiryayeva@gmail.com

В работе представлены результаты разработки схемы функционирования нейросетевой системы управления с использованием иммунных алгоритмов. На основе использования результатов нечёткой настройки регулятора получены результаты моделирования в среде MATLAB нейросетевой системы управления. В среде программирования MATLAB, с использованием пакетов Fuzzy Logic Toolbox, NNTool и Anfis сформирован алгоритм создания нейро-нечёткой сети для искусственной иммунной системы. Использована нечеткая многослойная нейронная сеть, в которой слои выполняют функции элементов системы нечеткого вывода. Получены результаты процесса обучения нейро-нечёткой сети в виде оптимальных значений системы.

Все переменные, и соответствующие им нечёткие множества, формируются на основе экспертного мнения специалистов. Выходные переменные описываются функциями системы нечеткого вывода Сугено. Алгоритм разработки модели гибридной сети для составления программного обеспечения в среде MATLAB осуществляется с помощью редактора ANFIS [1].

На основе методов нечеткой логики удастся спроектировать систему, способную эффективно функционировать в условиях наличия информации об объекте управления лишь качественного характера.

Для разработанной схемы моделирования стохастического оптимального управления для терапии сульфаниламидами в среде MATLAB Simulink [1], в ходе разработки нейро-нечёткой сети с экспертной базой, получено такое управляющее устройство, как блок нейро-нечёткого вывода, NF (рис. 1).

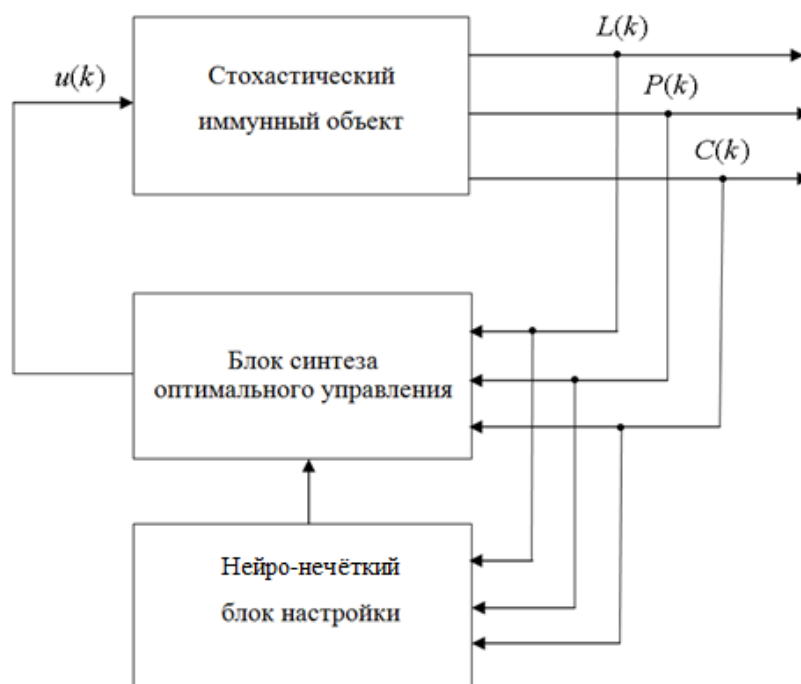


Рис. 1. Функциональная схема стохастической оптимальной иммунной системы управления с нейро-нечёткой настройкой

Функциональная схема нейро-нечеткой иммунной системы оптимального управления, в этом случае, имеет вид системы управления, блок оптимального управления которой дополнен нейро-нечётким блоком настройки (рис. 1). Данная схема допускает подстройку и обучение блока настройки. В этом случае система управления, использующая аппарат искусственных нейронных сетей и нечеткую логику, позволяет проводить идентификацию объекта и синтезировать для него оптимальный ответ, что дает возможность решать рассматриваемую задачу синтеза системы в условиях неопределенности на основе имеющихся экспериментальных данных, полученных на объекте. Применение нейро-нечёткой настройки в задаче синтеза системы терапии имеет существенный недостаток: информацию об объекте управления нейронная сеть получает в процессе обучения, а для этого необходим большой объем экспериментальных данных.

Блоки, включённые в схему в среде MATLAB, представляют собой элементы, связанные с уравнениями модели, разработанными в [2]:

- Plant: иммунная математическая модель объекта с выходными значениями L , P и C ;
- Stochastic optimal control (SOC): оптимальное управление, формирующее входной сигнал для иммунного объекта, u ;
- Neuro-fuzzy logic block: нейро-нечёткий блок подстройки с программным обеспечением `gibrid_medicine.fis`.

Разработанный программный продукт на основе редактора ANFIS среды MATLAB, в совокупности с математической моделью оптимального управления иммунным объектом, позволяет получить управление с нейро-нечёткой настройкой оптимального регулятора (рис. 2) [3].

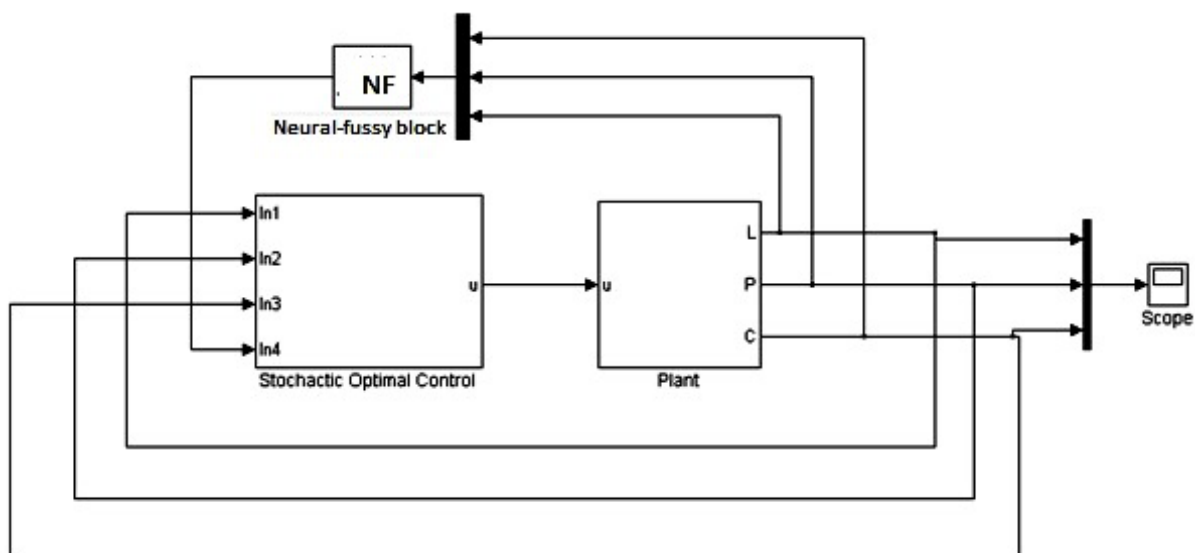


Рис. 2. Схема функционирования нейросетевой системы управления с иммунным объектом в среде MATLAB

В соответствии с данными блоками нечеткое управляющее устройство

вырабатывает управление, NF, на основе нечеткого множества управляющих воздействий с целью достижения критерия качества [2]. Вся исходная информация о стратегии управления хранится в базах функций принадлежности, и базе правил условного логического вывода. Нечеткому множеству управляющих воздействий соответствует функция принадлежности, которая формируется на основе выбранного критерия качества и ограничений в блоке SOC.

Математическая модель иммунной системы, представленная блоком Plant, описывается дискретными уравнениями, разработанными в [2]:

$$\begin{cases} L(k+1) = (-\gamma - f * P(k) - g * C(k)) * L(k) * T + L(k) \\ P(k+1) = (-\beta + \lambda * (1 / L(k)) - \rho * C(k)) * P(k) * T + P(k) \\ C(k+1) = (-\alpha + \eta * P(k)) * C(k) * T + C(k) + r * u(k) \end{cases} \quad (3.1)$$

где параметры и переменные:

L – количественное представление ресурсов иммунной системы (данная переменная включается в последующие два уравнения, как источник для роста соответствующих значений переменных P и C);

P – изменение выходных сигналов иммунной системы в зависимости от времени (данная переменная в иммунных моделях поддерживается за счёт ресурсов L , но уменьшается за счёт переменной C);

C – переменная, за счёт которой иммунная система регулирует изменение количества P (в иммунной системе данная переменная играет роль настраивающего сигнала, позволяя настраивать регулирование переменной P);

γ – темп естественного прироста ресурсов организма;

f – количество ресурсов иммунной системы, предназначенное для P ;

g – количество ресурсов иммунной системы, оценивает побочное влияние C ;

β – темп изменения количества P при отсутствии ресурсов организма;

α – коэффициент, связанный с темпом изменения количества C ;

η – терапевтическая доза сульфаниламидов, с учетом длительности выведения их из организма;

u – управляющее воздействие, сформированное на основе требований к динамике изменения P и C ;

k – дискретное время.

Разработанный программный продукт на основе редактора ANFIS среды MATLAB, в совокупности с математической моделью оптимального управления иммунным объектом, позволяет получить управление с нейро-нечёткой настройкой оптимального регулятора. Разработанный программный продукт отличается следующими характеристиками [4]:

Надежность – надежность функционирования программного продукта нейро-нечёткого вывода определяется работой без сбоев и стабильностью в работе программ, правильностью выполнения предписанных команд при обработке данных, возможностью диагностики возникающих ошибок в процессе работы приложения.

Эффективность – эффективность программного продукта нейро-нечёткого вывода, рассматривается с позиций прямого его назначения – требований пользователя и с точки зрения расхода вычислительных ресурсов, необходимых для его эксплуатации. Несмотря на сложность решаемых задач приложением, оно потребляет относительно малое количество оперативной памяти.

Учет человеческого фактора означает обеспечение интуитивно понятного интерфейса для работы фактического пользователя, наличие вспомогательного действия при допущении ошибки или обучающего модуля в составе программного средства. Нейро-нечёткую сеть иммунной системы можно настраивать в процессе работы.

Модифицируемость программного продукта нейро-нечёткого вывода представляет собой совокупность свойств, позволяющих вносить требуемые изменения. Система реализована таким образом, что с помощью инженера по знаниям, эксперты могут изменять, пополнять и удалять любые данные, включая те, которые расширяют функционал приложения. Система обладает удобным редактором базы знаний, нечетких и лингвистических переменных, а также других данных, влияющих на диалоговые формы.

Разработанное программное обеспечение должно использоваться в функциональной схеме выработки оптимального ответа на основе синтеза оптимального управления стохастическим иммунным объектом с нейро-нечёткой настройкой оптимального регулятора, на основе методов статистического анализа, динамического программирования Беллмана, генетического алгоритма, нечёткой логики

Список литературы

1. Ширяева О. И. Принципы построения нейро-нечёткой системы терапии лекарственными средствами / О. И. Ширяева // Материалы XXV Всероссийского семинара «Нейроинформатика, её приложения и анализ данных». – Красноярск. — 2017. – С. 112 – 118.
2. Ширяева О.И., Денисова Т.Г. Разработка искусственной иммунной системы оптимального управления терапевтическими дозами сульфаниламидов на основе нечеткой логики // Новосибирск: Проблемы информатики. — 2016. — № 2. С. 60 – 70.
3. Suparta W., Alhasa K.M. A comparison of ANFIS and MLP models for the prediction of perceptible water vapor // IEEE international conference on space science and communication (IconSpace). — 2013. — pp. 243 – 248.
4. Ерёмина В.В., Горожанина Ю.А. Проектирование экспертной системы диагностики на базе нечеткой логики // Современные научные исследования и инновации. — 2017. — № 6. E-print: <http://web.snauka.ru/issues/2017/06/83588>

О ЗАДАЧЕ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ РАБОТЫ СЕКВЕНАТОРА NGS

В.В. Шаров³, В.А. Кобец², Д.А. Кузьмин³, Г.А. Ходос², М.Г. Садовский^{1,2}

¹Институт вычислительного моделирования СО РАН,

²Сибирский федеральный университет, ИФБиТ,

³Сибирский федеральный университет, ИКИТ

*sharvadim07@ya.ru, victory.kobets@gmail.com, dm.kuzmin@gmail.com,
kalcifer@list.ru, msad@icm.krasn.ru*

Введение

Развитие современных методов секвенирования нуклеотидных последовательностей ставит перед исследователями ряд задач, связанных с изучением особенностей этого процесса, обеспеченного сложными современными технологиями. В грубом приближении секвенирование состоит в копировании мелких фрагментов исходного генома, при этом число копий и их расположение по геному таково, что возникает теоретическая возможность собрать всю последовательность целиком, используя информацию лишь об общих частях этих малых фрагментов.

Не обсуждая здесь всех важных деталей (как «мокрого» протокола, так и алгоритмических проблем сборки из огромного ансамбля фрагментов единого целого — например, единственности такой сборки) укажем одну важную проблему, на решение которой направлена наша работа. Ассемблирование (сборка полной последовательности из ансамбля коротких фрагментов) может быть осуществлена разными методами и разными алгоритмами, но большинство из них основываются на использовании графов Де Брюина [1]. Для того, чтобы такая сборка могла быть осуществлена, необходимо, чтобы фрагменты малой длины были скопированы вполне определённым образом. А именно, ключевым показателем качества секвенирования является так называемая глубина покрытия.

Глубина покрытия — это среднее число малых фрагментов (называемых ридами), которые покрывают каждый нуклеотид (символ в последовательности). Понятно, что эта величина не обязана быть постоянной для всех нуклеотидов последовательности; тем самым, разумным представляется введение локального показателя $H_L(n)$ глубины покрытия, где n — номер текущего нуклеотида в исходной последовательности. Очевидно, что

$$\bar{H} = N^{-1} \sum_{n=1}^N H_L(n) \quad (1)$$

является средней (по геному) глубиной покрытия; L здесь — это длина того участка нуклеотидной последовательности, по которому определяется локальная глубина покрытия. Качество секвенирования характеризуется двумя показателями: средней глубиной покрытия (1) и дисперсией набора значений локальной глубины покрытия $D\{H_L(n)\}$, определяемой по всему геному. Заметим, что в определении локальной глубины покрытия участвует не определяемый формально параметр — длина участка, по которому она определяется.

При этом фактически все алгоритмы (и их реализации) сборки целого генома по ансамблю ридов неявно опираются на предположение о том, что величина $D\{H_L(n)\}$ мала, и локальная глубина покрытия является весьма равномерной функцией n . Данное допущение совершенно не подтверждается конкретными биологическими данными: напротив, как правило, локальная глубина покрытия является очень и очень неоднородной по геному характеристикой. Существующие алгоритмы сборки способны работать и с весьма неоднородными по геному ансамблями ридов, однако и качество

сборки, и скорость работы алгоритма при этом могут заметно падать, не говоря уже о проблеме единственности сборки.

Данная работа направлена на изучение характера распределения ридов вдоль по геному и построение соответствующей имитационной модели работы секвенатора. Основной целью такого рода исследования является выявление и описание неоднородностей в распределении ридов вдоль по геному с тем, чтобы оценить их вклад в сложность задачи ассемблирования либо иной обработки такой генетической информации.

Результаты и обсуждение

Начнём представление результатов с анализа распределения локальной глубины покрытия (1) вдоль по геному некоторых организмов. Мы изучали бактериальные геномы, депонированные в EMBL-банке и NCBI банке геномных данных. Для анализа брались лишь те геномы, для которых также был доступен ансамбль ридов, полученных в результате секвенирования этого генома. В целях выполнения всех исследований, представленных в настоящей статье, со всеми геномами была выполнена следующая процедура (обратное картирование): на собранный и аннотированный геном выравнивались риды из того ансамбля, по которому этот геном ассемблировался, при этом использовались стандартные установки программы BowTie2, которая осуществляла выравнивание ридов на геном. Иными словами, допускалось как точное выравнивание (полное посимвольное совпадение рида и фрагмента в геноме), так и частичное — с некоторыми несовпадениями, уровень которых был невысок.

В ходе такого обратного картирования фиксировалась информация о том, в каком именно месте (номер нуклеотида) выравнивался данный рид; если рассматриваемый рид выравнивался в нескольких позициях вдоль по геному, то фиксировались все такие позиции. Затем выбиралась длина участка L (обычно в наших вычислительных экспериментах бралось $L \cong 10^3$ нуклеотидов) и шаг смещения T , с которым такие участки выделялись в геноме. Затем для каждого выделенного участка подсчитывалось число стартов ридов, попадавших в данный участок, что и давало величину локальной глубины покрытия (1).

На рис. 1 показаны некоторые типичные картины распределения локальной глубины покрытия (1) по геному для разных организмов

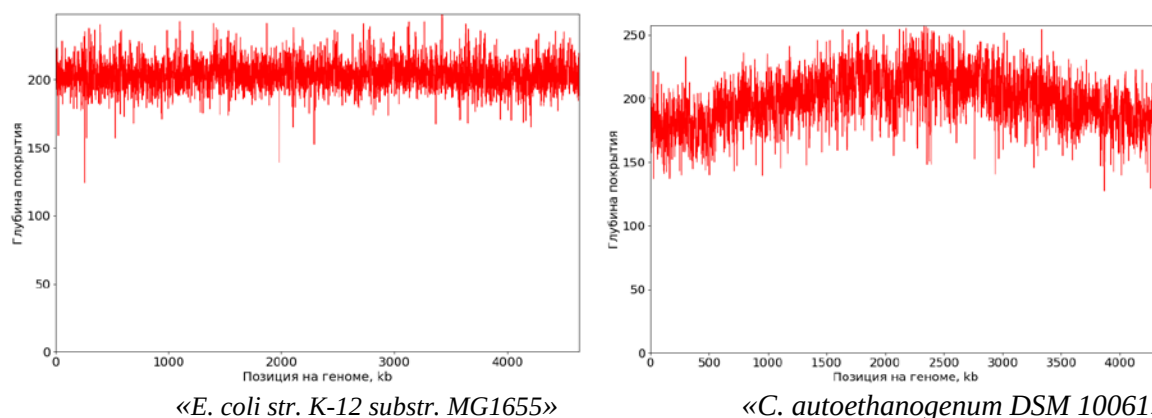


Рис. 1. Распределение локальной глубины покрытия вдоль по геномам *Escherichia coli* str. K-12 substr. MG1655 и *Clostridium autoethanogenum* DSM 10061, значения параметров $L = 1000$ и $T=1000$.

Локальная глубина покрытия (1) является показателем, характеризующим некоторый участок генома; понятно, что выбор величины L может существенно сказываться на наблюдаемой картине: если L будет достаточно велико, то

распределение будет «почти гладким» и многие важные детали станут недоступны для исследователя. Подход, основанный на изучении характера распределения точек стартов ридов вдоль по геному свободен от указанного недостатка.

Опишем постановку задачи подробнее. Пусть у нас имеется геном и соответствующий ему ансамбль ридов, которые картированы на геном. Пусть в результате картирования для каждого рида зафиксирован номер стартового нуклеотида. Здесь следует подчеркнуть, что в геномах существуют точки множественного старта ридов: в этих точках начинается не один, а несколько ридов. Очевидно, что чем больше число ридов, берущих начало в такой точке, тем меньше таких точек в геноме. Анализ характера распределения стартовых точек по числу ридов, берущих в них начало, представлен ниже.

Таблица 1

Значения условной энтропии для бинаризованного генома $\mathfrak{Z}$; список видов включает в себя: 1) *Acinetobacter baumannii* str. AB030; 2) *Campylobacter jejuni* subsp. *jejuni* NCTC 11168; 3) *Clostridium autoethanogenum* DSM 10061; 4) *Coprothermobacter proteolyticus* DSM 5265; 5) *Escherichia coli* str. K-12 substr. MG1655; 6) *Escherichia coli* O157:H7 str. Sakai; 7) *Enterobacter cloacae* subsp. *cloacae* ATCC 13047; 8) *Enterococcus faecalis* OG1RF; 9) *Mycobacterium tuberculosis* H37Rv; 10) *Saccharopolyspora erythraea* NRRL2338; 11) *Saccharomyces cerevisiae* S288C; 12) *Staphylococcus aureus* subsp. *aureus* NCTC 8325; 13) *Stanieria* sp. NIES-3757; 14) *Streptomyces* sp. Mg1; 15) *Yersinia pseudotuberculosis* YPIII

Вид	2	3	4	5	6	7	8	9	10
1	1,3110	1,2879	1,2631	1,2342	1,1915	1,1752	1,1544	1,1272	1,1157
2	1,3535	1,3322	1,3157	1,3149	1,3213	1,2800	1,2695	1,2423	1,2550
3	1,3088	1,3081	1,3093	1,3073	1,3110	1,3076	1,3067	1,3059	1,3251
4	1,0211	0,5888	0,4193	0,4475	0,5262	0,4673	0,5427	0,4851	0,5072
5	0,3797	0,3072	0,2819	0,2650	0,2682	0,2631	0,2833	0,2503	0,2776
6	1,2414	1,2411	1,2441	1,2375	1,2390	1,2433	1,2482	1,2232	1,2431
7	0,9216	0,7950	0,7200	0,6692	0,6367	0,6188	0,6272	0,5934	0,6409
8	1,3885	1,3799	1,3804	1,3748	1,3679	1,3615	1,3689	1,3541	1,3478
9	0,0011	0,0004	0,0005	0,0005	-0,0001	0,0007	0,0008	-0,0006	0,0009
10	1,2101	1,2098	1,2102	1,2084	1,2111	1,2088	1,2071	1,2060	1,2056
11	0,6470	0,6272	0,5863	0,5800	0,6014	0,5613	0,6010	0,5724	0,5732
12	1,3792	1,3597	1,3340	1,2839	1,2614	1,2270	1,2132	1,2195	1,2129
13	0,4889	0,4888	0,4888	0,4887	0,4891	0,4883	0,4892	0,4887	0,4873
14	1,0236	1,0234	1,0232	1,0230	1,0226	1,0231	1,0229	1,0233	1,0198
15	1,3837	1,3814	1,3820	1,3801	1,3800	1,3815	1,3766	1,3816	1,3694

Преобразуем теперь исходный геном в символьную последовательность $\mathfrak{Z}$. На первом этапе проигнорируем тот факт, что в некоторых позициях может начаться не один рид, а несколько. Преобразование исходного генома состоит в следующем: каждый нуклеотид в геноме заменим символом 0 (ноль), если в нём не стартует ни один рид; наоборот, если в данном нуклеотиде (позиции) в геноме стартует один или более ридов, заменим единицей. Тем самым, получим символьную последовательность из алфавита $\aleph = \{0,1\}$, полностью характеризующую распределение ридов вдоль по геному.

Ключевым вопросом здесь является характер чередований нулей и единиц. Для изучения этого распределения использовалось два подхода. Первый основывается на изучении распределения длин трактов. Любую подпоследовательность, выделяемую в пределах $\mathfrak{Z}$ и состоящую только из одних нулей или из одних единиц, будем называть **трактом**. Тем самым, вопрос о виде распределения стартовых точек вдоль по геному сводится к вопросу о виде распределения длин трактов.

Второй подход можно назвать микроскопическим, либо мезомасштабным. Он основан на изучении частотных и информационных характеристик ансамбля частотных словарей коротких слов (длиной q , $1 \leq q \leq 10$), формально выделяемых в $\mathfrak{Z}$. Частотным словарём W_q (толщины q) назовём список всех слов $\omega \in W_q$ этой длины, встречающихся в $\mathfrak{Z}$; словом будем называть связную последовательность ω из q стоящих подряд символов, встречающуюся в $\mathfrak{Z}$ [1]. Такое представление генома позволяет применить к нему весь арсенал средств статистического анализа [2, 3].

Один из способов поиска и выявления структурированности в символьных последовательностях состоит в оценке ожидаемой частоты слов заданной длины q , получаемой из частот слов меньшей длины (например, $q-1$). Опуская технические детали, которые могут быть найдены в [4 – 6], приведём саму формулу ожидаемой частоты $\tilde{f}_{t_1 t_2 t_3 \dots t_{q-1} t_q}$ слова длины q :

$$\tilde{f}_{t_1 t_2 t_3 \dots t_{q-1} t_q} = \frac{f_{t_1 t_2 t_3 \dots t_{q-1}} \times f_{t_2 t_3 \dots t_{q-1} t_q}}{f_{t_2 t_3 \dots t_{q-1}}}. \quad (2)$$

Одним из важных показателей упорядоченности является различие реальной частоты слова $f_{t_1 t_2 t_3 \dots t_{q-1} t_q}$ и его ожидаемой частоты $\tilde{f}_{t_1 t_2 t_3 \dots t_{q-1} t_q}$; существует много способов сравнения этих частот, мы будем пользоваться условной энтропией [4 – 6]. Опять же, опуская технические детали, приведём сразу окончательную формулу условной энтропии $\bar{S}_q$

$$\bar{S}_q = S_q - 2S_{q-1} + S_{q-2} \quad \text{и} \quad \bar{S}_2 = S_2 - 2S_1 \quad \text{для } q = 2. \quad (3)$$

В Таблице 1 представлены результаты определения условной энтропии (3) для 15 различных геномов, преобразованных в $\mathfrak{Z}$.

Обращают на себя внимание существенные различия в поседении условной энтропии (3), наблюдаемые для различных геномов. Для бактерии *A. baumannii str. AB030* наблюдается фактически линейное падение значений условной энтропии (3) по мере роста длины слов q . Уникальной представляется в этом наборе бактерия *Mycobacterium tuberculosis H37Rv*, для которой наблюдаются очень близкие значения условной энтропии (3), фактически не изменяющиеся по мере роста длины слов q . Впрочем, одна из возможных причин наблюдаемого эффекта — низкое качество ансамбля ридов. У всех остальных геномов (включая единственный эукариотический геном дрожжей) поведение условной энтропии (3) носит немонотонный характер с явно выраженной регулярностью.

Заключение

Самые первые результаты, полученные предложенным выше способом анализа выявления упорядоченности в чередованиях точек стартов ридов, показывают заметную эффективность этого метода. Тем не менее, анализ энтропийных характеристик не является исчерпывающим: энтропия характеризует лишь ансамбль целиком, в то время, как важны знания об индивидуальном распределении вполне конкретных слов.

Здесь важную роль должно сыграть изучение (имитационное моделирование, приближение и аналогии) характера распределения траншей нулей и единиц в $\mathfrak{Z}$. Сама по себе задача распадается на несколько связанных, но самостоятельных подзадач. Первая — собственно изучение распределения длины траншей. Вторая задача во многом аналогична первой, но для её решения требуется модификация последовательности $\mathfrak{Z}$: теперь необходим учёт числа стартов ридов, которые наблюдаются в каждом конкретном нуклеотиде исходной генетической последовательности. Иными словами, требуется расширение алфавита для $\mathfrak{Z}$, в котором бы каждый символ (на практике это может быть число) указывал на число стартов. Заметим, что на этом пути возможны некоторые технические трудности, обусловленные как заметно возросшей мощностью алфавита, так и очень малой частотой встречаемости некоторых символов и слов, в которых они содержатся. Нельзя не отметить здесь и повторение того же подхода, что и представленный выше — с вычислением условной энтропии (3) для по-новому определённой последовательности $\mathfrak{Z}$.

Ещё одно важное направление — поиск тех точек в $\mathfrak{Z}$, которые определяют своего рода сайты максимальной неопределённости на всех масштабах длин. Вернёмся к определению ожидаемой частоты (2); зафиксируем длину слов q и выделим в отдельный список все те слова, для которых реальная частота заметно отличается (в большую или меньшую сторону) от реальной. Пусть установлен некоторый порог отличия α и по нему выбран список таких слов; будем, следуя [5], называть информационно значимыми. Особенность такого рода слов состоит в том, что свойство быть информационно значимым не является монотонным по длине: некоторое слово длины q , являясь информационно значимым, вовсе не обязано входить в какое-либо информационно значимое слово длины $q+1$, либо содержать в себе информационно значимое слово длины $q-1$. Тем не менее, интересным представляется выделение всех таких слов, которые, являясь информационно значимыми каждой на своей длине, вложены одно в другое (более короткие сложены в более длинные) и построить монотонные цепочки таких вложений. Будем называть самые длинные такие слова сайтами неожиданности и рассмотрим их распределение вдоль по геному. Можно ожидать, что их распределение будет во многом отражать особенности работы секвенатора с реальными нуклеотидными последовательностями.

Список литературы

1. Pevzner P.A. Computational Molecular Biology. An Algorithmic Approach // MIT Press, 2001.
2. Bugaenko N. N., Gorban A.N., Sadovsky M. G. (1998) Maximum entropy method in analysis of genetic text and measurement of its information content // Open Systems & Information Dyn. 5(2): 265-278.
3. Садовский М.Г., Сенашова М.Ю., Путинцева Ю.А. Eight Clusters, Synchrony of Evolution and Unique Symmetry in Chloroplast Genomes: The Offering from Triplets. В: Chloroplasts and Cytoplasm: Structure and Functions (ser. Cell Biology Research Progress), Eds. Carl DeJesus and Lourdes Trask, 2018, Nova Science Publishers, Inc., N.-Y., pp.25 – 95.

4. Садовский М.Г. Comparison of real frequencies of strings vs. the expected ones reveals the information capacity of macromolecules // Journal of Biological Physics, 2003, vol. 29, № 1, pp. 23 – 38.
5. Sadovsky M.G., Shchepanovsky A., Putintzeva J. Genes, Information and Sense: Complexity and Knowledge Retrieval // Theory in Biosciences, 2008, vol. 127, pp. 69 – 78.
6. Sadovsky M.G. Information capacity of nucleotide sequences and its applications // Bulletin of Mathematical Biology, 2006, vol. 68, № 2, pp. 156 – 178.
7. Садовский М.Г., Сенашова М.Ю., Малышев А.В. Восьмикластерная структура геномов хлоропластов наземных растений // ЖОБ. 2018, т.79(2), С.124 – 134.
8. Садовский М.Г., Сенашова М.Ю., Малышев А.В. Chloroplast Genomes Exhibit Eight-Cluster Structuredness and Mirror Symmetry // LNBI, vol.10813, 2018, pp.186 – 196.

СПИСОК ДОКЛАДОВ

Али Яхья Сенан Аль-Баред, В.В. Жилин, Д.С. Кондрашов НЕЙРОННЫЕ СЕТИ ВСТРЕЧНОГО РАСПРОСТРАНЕНИЯ В СИСТЕМЕ УПРАВЛЕНИЯ ЭКЗОСКЕЛЕТОМ ПО СХЕМЕ СПЕЦИАЛИЗИРОВАННОГО ОБУЧЕНИЯ	3
М.М. Алкезуини, В.И. Горбаченко РЕШЕНИЕ КРАЕВЫХ ЗАДАЧ НА СЕТЯХ РАДИАЛЬНЫХ БАЗИСНЫХ ФУНКЦИЙ	9
Н.М.Богатов, А.Н.Богатова, С.А.Сухих, В.В.Филимонов СНИЖЕНИЕ НАПРЯЖЕНИЯ ФУНКЦИОНАЛЬНЫХ СИСТЕМ ОРГАНИЗМА С ПОМОЩЬЮ САМОРЕГУЛЯЦИИ	16
О.С.Володько, Л.А.Компаниец, Л.В.Гаврилова МЕТОД ГЛАВНЫХ КОМПОНЕНТ В ЗАДАЧЕ АНАЛИЗА ГИДРОФИЗИЧЕСКИХ ХАРАКТЕРИСТИК ОТКРЫТЫХ ВОДОЕМОВ	23
М.А. Габова, М.В. Гунер, О.И. Пятковский, Т.Ю. Солодовникова ПОСТРОЕНИЕ МОДЕЛИ ОЦЕНКИ НАУЧНОГО ПОТЕНЦИАЛА ИННОВАЦИОННОГО ПРОЕКТА НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ	30
И.К. Горбань, В.С. Густов, М.Ю. Сенашова, М.Г. Садовский НЕКОДИРУЮЩИЕ ОБЛАСТИ ГЕНОМОВ ХЛОРОПЛАСТОВ ОТРАЖАЮТ ТАКСОНОМИЮ ИХ НОСИТЕЛЕЙ	35
М.В. Гунер, Р.Е. Ткаченко, О.И. Пятковский РАЗРАБОТКА МОДЕЛЕЙ И СИСТЕМЫ ОЦЕНКИ И ПРОГНОЗА СПРОСА НА ПАРФЮМЕРНУЮ ПРОДУКЦИЮ НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ	54
А.В. Ершов, Ф.П. Капсаргин, К.В. Симонов НЕЙРОСЕТЕВАЯ ОЦЕНКА ВЫБОРА МЕТОДА ЛЕЧЕНИЯ МКБ	63
А.И.Захарова РЕЗУЛЬТАТ ЭКСПЕРИМЕНТА ПО РАБОТЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ПОДБОРУ ПРИПУСКОВ ЗАГОТОВОК	68
Н.А. Звеков, Н.В. Смыкова, О.И. Пятковский РАЗРАБОТКА АИС СОДЕЙСТВИЯ ТРУДОУСТРОЙСТВУ ВЫПУСКНИКОВ С ФУНКЦИЕЙ ОЦЕНКИ ПРОФПРИГОДНОСТИ НА ОСНОВЕ ГИБРИДНОЙ ЭКСПЕРТНОЙ СИСТЕМЫ	75

А.В. Киселев, Т.В. Петрова, О.В. Шаталова ГИБРИДНЫЕ РЕШАЮЩИЕ МОДУЛИ С ВИРТУАЛЬНЫМИ ПОТОКАМИ В КЛАССИФИКАТОРАХ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ СЛОЖНЫХ СИСТЕМ	79
И.О.Лимарев, С.А.Субботин, А.А.Олейник, И.В.Дрокин МЕТОДЫ ПРЕОБРАЗОВАНИЯ СИГНАЛОВ БОЛЬШОЙ РАЗМЕРНОСТИ В ДИАГНОСТИЧЕСКИХ МОДЕЛЯХ УПРАВЛЯЕМЫХ ДАННЫМИ	86
С.В. Листопад МОДЕЛИРОВАНИЕ ДИВЕРГЕНТНОГО МЫШЛЕНИЯ В МАЛЫХ КОЛЛЕКТИВАХ ЭКСПЕРТОВ ГИБРИДНЫМИ ИНТЕЛЛЕКТУАЛЬНЫМИ МНОГОАГЕНТНЫМИ СИСТЕМАМИ	93
Т.В. Петрова, С.А. Филист, В.В. Жилин УНИВЕРСАЛЬНЫЙ АППРОКСИМАТОР НА ОСНОВЕ МГУА- МОДЕЛЕЙ ДЛЯ ФОРМИРОВАНИЯ ВИРТУАЛЬНЫХ ПОТОКОВ	99
К.В. Пушкарев ГЛОБАЛЬНАЯ ОПТИМИЗАЦИЯ НА ОСНОВЕ НЕЙРОСЕТЕВОЙ АППРОКСИМАЦИИ ИНВЕРСНЫХ ЗАВИСИМОСТЕЙ С ЭВОЛЮЦИОННОЙ НАСТРОЙКОЙ ПАРАМЕТРОВ	106
Г.А. Самигулина, Ж.С. Лукманова ДИСТАНЦИОННОЕ ОБУЧЕНИЕ ЛЮДЕЙ С ОСЛАБЛЕННЫМ ЗРЕНИЕМ НА ОСНОВЕ КОГНИТИВНОЙ SMART – ТЕХНОЛОГИИ	112
В.П.Сизиков, В.И. Разумо ЗАПОЛНЕНИЕ ПУСТОТ СРЕДСТВАМИ ДИС-ТЕХНОЛОГИИ	117
Травкина Е.С., В.А.Углев, Ступницкий М.С. ОБРАБОТКА БАЗЫ ЗНАНИЙ В СИСТЕМЕ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ИНСТРУКТОРОВ ПРИ ВЫБОРЕ ПРОГРАММ ТРЕНИРОВКИ В ЛЮБИТЕЛЬСКОМ СПОРТЕ	123
В.А.Углев КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ОРГАНИЗАЦИИ ДАННЫХ РЕЕСТРА ФРАГМЕНТИРОВАННЫХ ЭЛЕКТРОННЫХ РЕСУРСОВ	128
В.А.Углев, Я.А. Феодоров ОЦЕНКА ПОКАЗАТЕЛЯ ДОЛИ АВТОРСКОГО ВКЛАДА, АВТОРСКОЙ «ЧИСТОТЫ» И ЦИТИРУЕМОСТИ ФРАГМЕНТИРОВАННЫХ РЕСУРСОВ	133
И.С.Фомин, С.Р.Орлова, А.В.Бахшиев ПРИМЕНЕНИЕ КОМБИНАЦИИ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ И КЛАССИФИКАЦИИ ЛЮДЕЙ НА ВИДЕОИЗОБРАЖЕНИИ	138

Р.А. Чиженкова АНАЛИЗ ИНИЦИАТИВНЫХ НАУЧНЫХ ПРОЕКТОВ ПО ОБЛАСТИ ЗНАНИЯ "ФИЗИКА, АСТРОНОМИЯ", ПОДДЕРЖАННЫХ РОССИЙСКИМ ФОНДОМ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ	144
О.И. Ширяева РАЗВИТИЕ ИММУННЫХ АЛГОРИТМОВ НА КЛАСС НЕЙРОСЕТЕВЫХ СИСТЕМ УПРАВЛЕНИЯ	149
В.В. Шаров, В.А. Кобец, Д.А. Кузьмин, Г.А. Ходос, М.Г. Садовский О ЗАДАЧЕ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ РАБОТЫ ЕКВЕНАТОРА NGS	153

СПИСОК АВТОРОВ

Ф.И.О.	Место работы	Стр.
Али Яхья Сенан Аль-Бареда	117198, Москва, ул. Миклухо-Маклая, 6, Российский Университет Дружбы Народов, кафедра кибернетики и мехатроники, E-mail: SFilist@gmail.com	3
Алкезуини Мухи Мартада	440026, г. Пенза, ул. Красная, д. 40. Пензенский государственный университет E-mail: mohieit@mail.ru	9
Бахшиев Александр Валерьевич	Центральный научно-исследовательский и опытно- конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК) E-mail: bakhshiev_av@spbstu.ru	138
Богатов Николай Маркович	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет», E-mail: bogatov@phys.kubsu.ru	16
Богатова Анна Николаевна	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет», т. 8(861)2199501266, E-mail: bogatov@phys.kubsu.ru	16
Володько Ольга Станиславовна	236022, Калининград, просп. Мира, 1-3, Институт океанологии АО РАН E-mail: olga.pitalskaya@gmail.com	23
Габова Мария Андреевна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: gabova_ma@mail.ru	30
Гаврилова Людмила Владимировна	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет E-mail: lv.gavrilova@gmail.com	23
Горбань Инна Константиновна	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет E-mail: inn.gorban@gmail.com	35
Горбаченко Владимир Иванович	440026, Пенза, ул. Красная, 40, Пензенский государственный университет E-mail: gorvi@mail.ru	9
Гунер Михаил Викторович	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: horyzont1@mail.ru	30, 54
Густов Владимир Сергеевич	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет E-mail: v.gustows@mail.ru	35
Дрокин И.В.	Центр испытаний УГК "Мотор Сич", отдел исследований E-mail: igor.dv.mail@gmail.com	86

Ершов Артем Владимирович	660022, Красноярск, ул. Партизана Железняка, 1, ИПО Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого E-mail: kapsargin@mail.ru	63
Жилин Валерий Вячеславович	305040, г. Курск, ул. 50 лет Октября, 94, Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет» E-mail: vvzhilin61@gmail.com	3, 99
Захарова Александра Ивановна	662971, Железногорск, ул. Кирова, 12-а, Железногорский филиал Сибирского федерального университета, АО «ИСС», г. Железногорск E-mail: alex69_5@mail.ru	68
Звеков Никита Александрович	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова E-mail: nzvekov@inbox.ru	75
Капсаргин Федор Петрович	660022, Красноярск, ул. Партизана Железняка, 1, ИПО Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого E-mail: kapsargin@mail.ru	63
Киселев Алексей Викторович	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный государственный университет E-mail: Kiselevalexey@gmail.com	79
Кобец Виктория Алексеевна	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет, ИФБиТ E-mail: victory.kobets@gmail.com	
Компаниец Лидия Алексеевна	660036, Красноярск, Академгородок, Институт вычислительного моделирования СО РАН E-mail: kla@icm.krasn.ru	23
Кондрашов Дмитрий Сергеевич	305040, г. Курск, ул. 50 лет Октября, 94, Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет» E-mail: kondrashov012@mail.ru	3
Кузьмин Дмитрий Александрович	660074, Красноярск, ул. Академика Киренского, д.26б, Сибирский федеральный университет, Институт космических и информационных технологий E-mail: dm.kuzmin@gmail.com	153
Лимарев Игорь Олегович	Украина, 69063, Запорожье, ул. Жуковского, 64, Запорожский национальный технический университет E-mail: limarevigor@gmail.com	86
Листопад Сергей Викторович	Калининград, ул. Гостиная, д.5, Калининградский филиал ФИЦ "Информатика и управление" РАН E-mail: ser-list-post@yandex.ru	93
Лукманова Жанар Сапарбековна	050013, Алматы, ул. Сатпаева 22, ГУК 703 каб, КазННТУ им. К.И. Сатпаева E-mail: azeshova@mail.ru	112
Олейник Андрей Александрович	Украина, 69063, г. Запорожье, ул. Жуковского, 64, Запорожский национальный технический университет	86

	E-mail: olejnikaa@gmail.com	
Орлова Светлана Ринатовна	Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК) E-mail: bakhshiev_av@spbstu.ru	138
Петрова Татьяна Владимировна	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный государственный университет E-mail: ptata@ya.ru	79, 99
Пушкарев Кирилл Владимирович	660074, Красноярск, ул. Академика Киренского, д.26б, Сибирский федеральный университет, Институт космических и информационных технологий, каф. Вычислительной техники E-mail: pushkaryov@gmail.com	106
Пятковский Олег Иванович	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова E-mail: poi1952@mail.ru	30, 54, 75
Разумов Владимир Ильич	Омский государственный университет им. Ф.М. Достоевского, Россия, 644077, Омск, проспект Мира 55-а E-mail: razumov@omsu.ru	117
Садовский Михаил Георгиевич	660036, Красноярск, Академгородок, Институт вычислительного моделирования СО РАН E-mail: msad@icm.krasn.ru	35, 153
Самигулина Галина Ахметовна	050010, Республика Казахстан, г. Алматы, ул. Пушкина, 125, Институт информационных и вычислительных технологий E-mail: galinasamigulina@mail.ru	112
Сенашова Мария Юрьевна	Институт вычислительного моделирования СО РАН, 660036, Красноярск, Академгородок E-mail: msen@icm.krasn.ru	35
Сизиков Виктор Петрович	Омский государственный университет путей сообщения, 644046, Россия, г. Омск, пр. Маркса, 35. E-mail: v_p_sizikov@mail.ru	117
Симонов Константин Васильевич	660036, Красноярск, Академгородок, Институт вычислительного моделирования СО РАН, т. 89135954902 E-mail: simonovkv@icm.krasn.ru	63
Смыкова Наталья Владимировна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова E-mail: n-smykova@mail.ru	75
Солодовникова Татьяна Юрьевна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: solodovnikova.2010@yandex.ru	30
Ступницкий Михаил Сергеевич	г. Железногорск, ул. Кирова, ФГБУЗ КБ №51 ФМБА России E-mail: stmishgun@gmail.com	123
Субботин Сергей Александрович	Украина, 69063, Запорожье, ул. Жуковского, 64, Запорожский национальный технический университет E-mail: subbotin@zntu.edu.ua	86
Сухих Станислав Алексеевич	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет»,	16

	E-mail: suhih_52@mail.ru	
Ткаченко Регина Евгеньевна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: horyzont1@mail.ru	54
Травкина Екатерина Сергеевна	Спортивно-танцевальная школа «Пурпур» г. Железногорск, Сибирский федеральный университет E-mail: katya_ml15@mail.ru	123
Углев Виктор Александрович	662971, Железногорск, ул. Кирова, 12-а, Железногорский филиал Сибирского федерального университета E-mail: uglev-v@yandex.ru	123, 128, 133
Фазилова Ксения Наильевна	119454, Россия, г. Москва, пр. Вернадского, 78, Московский технологический университет (МИРЭА) E-mail: fazilova@mirea.ru	
Федоров Ян Алексеевич	662971, Железногорск, ул. Кирова, 12-а, Железногорский филиал Сибирского федерального университета E-mail: feodorov.yan@mail.ru	133
Филимонов Вадим Владимирович	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет», E-mail: vadik2393@gmail.com	16
Филист Сергей Алексеевич	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный государственный университет E-mail: Sfilist@gmail.com	99
Фомин Иван Сергеевич	Центральный научно-исследовательский и опытно- конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК) E-mail: i.fomin@rtc.ru	138
Ходос Георгий Александрович	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет, ИФБиТ E-mail: kalcifer@list.ru	153
Чиженкова Рогнеда Александровна	142290, г. Пушкино Московской области, Институт биофизики клетки РАН E-mail: chizhenkova@mail.ru	144
Шаров Вадим Витальевич	660074, Красноярск, ул. Академика Киренского, д.26б, Сибирский федеральный университет, Институт космических и информационных технологий E-mail: sharvadim07@ya.ru	153
Шаталова Ольга Владимировна	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный государственный университет E-mail: shatolg@mail.ru	79
Ширяева Ольга Ивановна	050010, Республика Казахстан, Алматы, ул. Пушкина, 125, Институт информационных и вычислительных технологий КН МОН РК E-mail: oshiryayeva@gmail.com	149

СПИСОК ДОКЛАДОВ

Али Яхья Сенан Аль-Баред, В.В. Жилин, Д.С. Кондрашов НЕЙРОННЫЕ СЕТИ ВСТРЕЧНОГО РАСПРОСТРАНЕНИЯ В СИСТЕМЕ УПРАВЛЕНИЯ ЭКЗОСКЕЛЕТОМ ПО СХЕМЕ СПЕЦИАЛИЗИРОВАННОГО ОБУЧЕНИЯ	3
М.М. Алкезуини, В.И. Горбаченко РЕШЕНИЕ КРАЕВЫХ ЗАДАЧ НА СЕТЯХ РАДИАЛЬНЫХ БАЗИСНЫХ ФУНКЦИЙ	9
Н.М.Богатов, А.Н.Богатова, С.А.Сухих, В.В.Филимонов СНИЖЕНИЕ НАПРЯЖЕНИЯ ФУНКЦИОНАЛЬНЫХ СИСТЕМ ОРГАНИЗМА С ПОМОЩЬЮ САМОРЕГУЛЯЦИИ	16
О.С.Володько, Л.А.Компаниец, Л.В.Гаврилова МЕТОД ГЛАВНЫХ КОМПОНЕНТ В ЗАДАЧЕ АНАЛИЗА ГИДРОФИЗИЧЕСКИХ ХАРАКТЕРИСТИК ОТКРЫТЫХ ВОДОЕМОВ	23
М.А. Габова, М.В. Гунер, О.И. Пятковский, Т.Ю. Солодовникова ПОСТРОЕНИЕ МОДЕЛИ ОЦЕНКИ НАУЧНОГО ПОТЕНЦИАЛА ИННОВАЦИОННОГО ПРОЕКТА НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ	30
И.К. Горбань, В.С. Густов, М.Ю. Сенашова, М.Г. Садовский НЕКОДИРУЮЩИЕ ОБЛАСТИ ГЕНОМОВ ХЛОРОПЛАСТОВ ОТРАЖАЮТ ТАКСОНОМИЮ ИХ НОСИТЕЛЕЙ	35
М.В. Гунер, Р.Е. Ткаченко, О.И. Пятковский РАЗРАБОТКА МОДЕЛЕЙ И СИСТЕМЫ ОЦЕНКИ И ПРОГНОЗА СПРОСА НА ПАРФЮМЕРНУЮ ПРОДУКЦИЮ НА ОСНОВЕ ПРИМЕНЕНИЯ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ	54
А.В. Ершов, Ф.П. Капсаргин, К.В. Симонов НЕЙРОСЕТЕВАЯ ОЦЕНКА ВЫБОРА МЕТОДА ЛЕЧЕНИЯ МКБ	63
А.И.Захарова РЕЗУЛЬТАТ ЭКСПЕРИМЕНТА ПО РАБОТЕ ИНТЕЛЛЕКТУАЛЬНОЙ СИСТЕМЫ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ПО ПОДБОРУ ПРИПУСКОВ ЗАГОТОВОК	68
Н.А. Звеков, Н.В. Смыкова, О.И. Пятковский	

РАЗРАБОТКА АИС СОДЕЙСТВИЯ ТРУДОУСТРОЙСТВУ ВЫПУСКНИКОВ С ФУНКЦИЕЙ ОЦЕНКИ ПРОФПРИГОДНОСТИ НА ОСНОВЕ ГИБРИДНОЙ ЭКСПЕРТНОЙ СИСТЕМЫ	75
А.В. Киселев, Т.В. Петрова, О.В. Шаталова ГИБРИДНЫЕ РЕШАЮЩИЕ МОДУЛИ С ВИРТУАЛЬНЫМИ ПОТОКАМИ В КЛАССИФИКАТОРАХ ФУНКЦИОНАЛЬНОГО СОСТОЯНИЯ СЛОЖНЫХ СИСТЕМ	79
И.О.Лимарев, С.А.Субботин, А.А.Олейник, И.В.Дрокин МЕТОДЫ ПРЕОБРАЗОВАНИЯ СИГНАЛОВ БОЛЬШОЙ РАЗМЕРНОСТИ В ДИАГНОСТИЧЕСКИХ МОДЕЛЯХ УПРАВЛЯЕМЫХ ДАННЫМИ	86
С.В. Листопад МОДЕЛИРОВАНИЕ ДИВЕРГЕНТНОГО МЫШЛЕНИЯ В МАЛЫХ КОЛЛЕКТИВАХ ЭКСПЕРТОВ ГИБРИДНЫМИ ИНТЕЛЛЕКТУАЛЬНЫМИ МНОГОАГЕНТНЫМИ СИСТЕМАМИ	93
Т.В. Петрова, С.А. Филист, В.В. Жилин УНИВЕРСАЛЬНЫЙ АППРОКСИМАТОР НА ОСНОВЕ МГУА- МОДЕЛЕЙ ДЛЯ ФОРМИРОВАНИЯ ВИРТУАЛЬНЫХ ПОТОКОВ	99
К.В. Пушкарев ГЛОБАЛЬНАЯ ОПТИМИЗАЦИЯ НА ОСНОВЕ НЕЙРОСЕТЕВОЙ АППРОКСИМАЦИИ ИНВЕРСНЫХ ЗАВИСИМОСТЕЙ С ЭВОЛЮЦИОННОЙ НАСТРОЙКОЙ ПАРАМЕТРОВ	106
Г.А. Самигулина, Ж.С. Лукманова ДИСТАНЦИОННОЕ ОБУЧЕНИЕ ЛЮДЕЙ С ОСЛАБЛЕННЫМ ЗРЕНИЕМ НА ОСНОВЕ КОГНИТИВНОЙ SMART – ТЕХНОЛОГИИ	112
В.П.Сизиков, В.И. Разумо ЗАПОЛНЕНИЕ ПУСТОТ СРЕДСТВАМИ ДИС-ТЕХНОЛОГИИ	117
Травкина Е.С., В.А.Углев, Ступницкий М.С. ОБРАБОТКА БАЗЫ ЗНАНИЙ В СИСТЕМЕ ПОДДЕРЖКИ ПРИНЯТИЯ РЕШЕНИЙ ИНСТРУКТОРОВ ПРИ ВЫБОРЕ ПРОГРАММ ТРЕНИРОВКИ В ЛЮБИТЕЛЬСКОМ СПОРТЕ	123
В.А.Углев КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ ОРГАНИЗАЦИИ ДАННЫХ РЕЕСТРА ФРАГМЕНТИРОВАННЫХ ЭЛЕКТРОННЫХ РЕСУРСОВ	128
В.А.Углев, Я.А. Феодоров	

ОЦЕНКА ПОКАЗАТЕЛЯ ДОЛИ АВТОРСКОГО ВКЛАДА, АВТОРСКОЙ «ЧИСТОТЫ» И ЦИТИРУЕМОСТИ ФРАГМЕНТИРОВАННЫХ РЕСУРСОВ	133
И.С.Фомин, С.Р.Орлова, А.В.Бахшиев ПРИМЕНЕНИЕ КОМБИНАЦИИ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ И КЛАССИФИКАЦИИ ЛЮДЕЙ НА ВИДЕОИЗОБРАЖЕНИИ	138
Р.А. Чиженкова АНАЛИЗ ИНИЦИАТИВНЫХ НАУЧНЫХ ПРОЕКТОВ ПО ОБЛАСТИ ЗНАНИЯ "ФИЗИКА, АСТРОНОМИЯ", ПОДДЕРЖАННЫХ РОССИЙСКИМ ФОНДОМ ФУНДАМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ	144
О.И. Ширяева РАЗВИТИЕ ИММУННЫХ АЛГОРИТМОВ НА КЛАСС НЕЙРОСЕТЕВЫХ СИСТЕМ УПРАВЛЕНИЯ	149
В.В. Шаров, В.А. Кобец, Д.А. Кузьмин, Г.А. Ходос, М.Г. Садовский О ЗАДАЧЕ ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ РАБОТЫ ЕКВЕНАТОРА NGS	153

СПИСОК АВТОРОВ

Ф.И.О.	Место работы	Стр.
Али Яхья Сенан Аль-Баред	117198, Москва, ул. Миклухо-Маклая, 6, Российский Университет Дружбы Народов, кафедра кибернетики и мехатроники, E-mail: SFilist@gmail.com	3
Алкезуини Мухи Мартада	440026, г. Пенза, ул. Красная, д. 40. Пензенский государственный университет E-mail: mohieit@mail.ru	9
Бахшиев Александр Валерьевич	Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК) E-mail: bakhshiev_av@spbstu.ru	138
Богатов Николай Маркович	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет», E-mail: bogatov@phys.kubsu.ru	16
Богатова Анна Николаевна	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет», т. 8(861)2199501266, E-mail: bogatov@phys.kubsu.ru	16
Володько Ольга Станиславовна	236022, Калининград, просп. Мира, 1-3, Институт океанологии АО РАН E-mail: olga.pitalskaya@gmail.com	23
Габова Мария Андреевна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: gabova_ma@mail.ru	30
Гаврилова Людмила Владимировна	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет E-mail: lv.gavrilova@gmail.com	23
Горбань Инна Константиновна	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет E-mail: inn.gorban@gmail.com	35
Горбаченко Владимир Иванович	440026, Пенза, ул. Красная, 40, Пензенский государственный университет E-mail: gorvi@mail.ru	9
Гунер Михаил Викторович	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: horyzont1@mail.ru	30, 54
Густов Владимир Сергеевич	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет E-mail: v.gustows@mail.ru	35
Дрокин И.В.	Центр испытаний УГК "Мотор Сич", отдел исследований E-mail: igor.dv.mail@gmail.com	86

Ершов Артем Владимирович	660022, Красноярск, ул. Партизана Железняка, 1, ИПО Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого E-mail: kapsargin@mail.ru	63
Жилин Валерий Вячеславович	305040, г. Курск, ул. 50 лет Октября, 94, Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет» E-mail: vvzhilin61@gmail.com	3, 99
Захарова Александра Ивановна	662971, Железногорск, ул. Кирова, 12-а, Железногорский филиал Сибирского федерального университета, АО «ИСС», г. Железногорск E-mail: alex69_5@mail.ru	68
Звеков Никита Александрович	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова E-mail: nzvekov@inbox.ru	75
Капсаргин Федор Петрович	660022, Красноярск, ул. Партизана Железняка, 1, ИПО Красноярский государственный медицинский университет имени профессора В.Ф. Войно-Ясенецкого E-mail: kapsargin@mail.ru	63
Киселев Алексей Викторович	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный государственный университет E-mail: Kiselevalexey@gmail.com	79
Кобец Виктория Алексеевна	660041, г. Красноярск, пр. Свободный, 79, Сибирский федеральный университет, ИФБиТ E-mail: victory.kobets@gmail.com	
Компаниец Лидия Алексеевна	660036, Красноярск, Академгородок, Институт вычислительного моделирования СО РАН E-mail: kla@icm.krasn.ru	23
Кондрашов Дмитрий Сергеевич	305040, г. Курск, ул. 50 лет Октября, 94, Федеральное государственное бюджетное образовательное учреждение высшего образования «Юго-Западный государственный университет» E-mail: kondrashov012@mail.ru	3
Кузьмин Дмитрий Александрович	660074, Красноярск, ул. Академика Киренского, д.26б, Сибирский федеральный университет, Институт космических и информационных технологий E-mail: dm.kuzmin@gmail.com	153
Лимарев Игорь Олегович	Украина, 69063, Запорожье, ул. Жуковского, 64, Запорожский национальный технический университет E-mail: limarevigor@gmail.com	86
Листопад Сергей Викторович	Калининград, ул. Гостиная, д.5, Калининградский филиал ФИЦ "Информатика и управление" РАН E-mail: ser-list-post@yandex.ru	93
Лукманова Жанар Сапарбековна	050013, Алматы, ул. Сатпаева 22, ГУК 703 каб, КазННТУ им. К.И. Сатпаева E-mail: azeshova@mail.ru	112
Олейник Андрей Александрович	Украина, 69063, г. Запорожье, ул. Жуковского, 64, Запорожский национальный технический университет E-mail: olejnikaa@gmail.com	86

Орлова Светлана Ринатовна	Центральный научно-исследовательский и опытно-конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК) E-mail: bakhshiev_av@spbstu.ru	138
Петрова Татьяна Владимировна	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный государственный университет E-mail: ptata@ya.ru	79, 99
Пушкарёв Кирилл Владимирович	660074, Красноярск, ул. Академика Киренского, д.266, Сибирский федеральный университет, Институт космических и информационных технологий, каф. Вычислительной техники E-mail: pushkaryov@gmail.com	106
Пятковский Олег Иванович	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова E-mail: poi1952@mail.ru	30, 54, 75
Разумов Владимир Ильич	Омский государственный университет им. Ф.М. Достоевского, Россия, 644077, Омск, проспект Мира 55-а E-mail: razumov@omsu.ru	117
Садовский Михаил Георгиевич	660036, Красноярск, Академгородок, Институт вычислительного моделирования СО РАН E-mail: msad@icm.krasn.ru	35, 153
Самигулина Галина Ахметовна	050010, Республика Казахстан, г. Алматы, ул. Пушкина, 125, Институт информационных и вычислительных технологий E-mail: galinasamigulina@mail.ru	112
Сенашова Мария Юрьевна	Институт вычислительного моделирования СО РАН, 660036, Красноярск, Академгородок E-mail: msen@icm.krasn.ru	35
Сизиков Виктор Петрович	Омский государственный университет путей сообщения, 644046, Россия, г. Омск, пр. Маркса, 35. E-mail: v_p_sizikov@mail.ru	117
Симонов Константин Васильевич	660036, Красноярск, Академгородок, Институт вычислительного моделирования СО РАН, т. 89135954902 E-mail: simonovkv@icm.krasn.ru	63
Смыкова Наталья Владимировна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова E-mail: n-smykova@mail.ru	75
Солодовникова Татьяна Юрьевна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: solodovnikova.2010@yandex.ru	30
Ступницкий Михаил Сергеевич	г. Железнодорожск, ул. Кирова, ФГБУЗ КБ №51 ФМБА России E-mail: stmishgun@gmail.com	123
Субботин Сергей Александрович	Украина, 69063, Запорожье, ул. Жуковского, 64, Запорожский национальный технический университет E-mail: subbotin@zntu.edu.ua	86
Сухих Станислав	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет»,	16

Алексеевич	E-mail: suhih_52@mail.ru	
Ткаченко Регина Евгеньевна	656038, Российская Федерация, Алтайский край, г. Барнаул, пр. Ленина, 46, Алтайский государственный технический университет им. И. И. Ползунова, E-mail: horyzont1@mail.ru	54
Травкина Екатерина Сергеевна	Спортивно-танцевальная школа «Пурпур» г. Железнодорожск, Сибирский федеральный университет E-mail: katya_ml15@mail.ru	123
Углев Виктор Алек- сандрович	662971, Железнодорожск, ул. Кирова, 12-а, Железнодорожский филиал Сибирского федерального университета E-mail: uglev-v@yandex.ru	123, 128, 133
Фазилова Ксения Наильевна	119454, Россия, г. Москва, пр. Вернадского, 78, Московский технологический университет (МИРЭА) E-mail: fazilova@mirea.ru	
Феодоров Ян Алек- сеев	662971, Железнодорожск, ул. Кирова, 12-а, Железнодорожский филиал Сибирского федерального университета E-mail: feodorov.yan@mail.ru	133
Филимонов Вадим Владимирович	350040, г. Краснодар, ул. Ставропольская, 149, ФГБОУ ВПО «Кубанский государственный университет», E-mail: vadik2393@gmail.com	16
Филист Сергей Алексеевич	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный гос- ударственный университет E-mail: Sfilist@gmail.com	99
Фомин Иван Серге- евич	Центральный научно-исследовательский и опытно- конструкторский институт робототехники и технической кибернетики (ЦНИИ РТК) E-mail: i.fomin@rtc.ru	138
Ходос Георгий Александрович	660041, г. Красноярск, пр. Свободный, 79, Сибирский федераль- ный университет, ИФБиТ E-mail: kalcifer@list.ru	153
Чиженкова Рогнеда Александровна	142290, г. Пушкино Московской области, Институт биофизи- ки клетки РАН E-mail: chizhenkova@mail.ru	144
Шаров Вадим Вита- льевич	660074, Красноярск, ул. Академика Киренского, д.26б, Си- бирский федеральный университет, Институт космических и информационных технологий E-mail: sharvadim07@ya.ru	153
Шаталова Ольга Владимировна	305040, г. Курск, ул. 50 лет Октября, 94, Юго-Западный гос- ударственный университет E-mail: shatolg@mail.ru	79
Ширяева Ольга Ивановна	050010, Республика Казахстан, Алматы, ул. Пушкина, 125, Институт информационных и вычислительных технологий КН МОН РК E-mail: oshiryayeva@gmail.com	149

Научное издание

Нейроинформатика, её приложения и анализ данных

Материалы XXVI Всероссийского семинара 28 сентября – 30 сентября 2018 года

Редактор М.Ю. Сенашова
Компьютерная верстка: М.Ю. Сенашова

Подписано в печать «15» августа 2018 г. Формат 60 × 90/8. (А4)
Бумага офсетная. Печать плоская.
Усл. печ. л. 11,0. Уч.-изд. л. 11,3.
Тираж 500 экз. Заказ 15

Отпечатано в типографии ИВМ СО РАН
660036, Красноярск, Академгородок