

На правах рукописи



СОБОЛЕВСКИЙ Владислав Алексеевич

**КОМПЛЕКСНАЯ АВТОМАТИЗАЦИЯ СИНТЕЗА ИСКУССТВЕННЫХ
НЕЙРОННЫХ СЕТЕЙ ПРЯМОГО РАСПРОСТРАНЕНИЯ**

Специальность 2.3.5 – Математическое и программное обеспечение
вычислительных систем, комплексов и компьютерных сетей

**Автореферат
диссертации на соискание учёной степени
кандидата технических наук**

Санкт-Петербург – 2022

Работа выполнена в Федеральном государственном бюджетном учреждении науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) в лаборатории информационных технологий в системном анализе и моделировании.

Научный руководитель:

СОКОЛОВ Борис Владимирович,

доктор технических наук, профессор, заслуженный деятель науки РФ, главный научный сотрудник, руководитель лаборатории информационных технологий в системном анализе и моделировании Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук».

Официальные оппоненты:

ЕФИМОВ Владимир Васильевич,

доктор технических наук, профессор, профессор кафедры автономных систем управления Федерального государственного бюджетного военного образовательного учреждения высшего образования «Военно-космическая академия имени А. Ф. Можайского».

СТАНКЕВИЧ Лев Александрович,

кандидат технических наук, доцент, доцент Высшей школы автоматизации и робототехники Института машиностроения, материалов и транспорта Федерального государственного автономного образовательного учреждения высшего образования «Санкт-Петербургский политехнический университет Петра Великого».

Ведущая организация

Федеральное государственное бюджетное учреждение науки «Институт проблем управления им. В. А. Трапезникова Российской академии наук».

Защита диссертации состоится «23» марта 2023 г. в 14 часа 00 минут на заседании диссертационного совета 24.1.206.01, созданного на базе Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) по адресу: 199178, Санкт-Петербург, 14-я линия В.О., 39, каб. 401, e-mail: dc@spcras.ru. Факс: (812) 328-44-50, тел: (812) 328-33-11.

С диссертацией и авторефератом можно ознакомиться в отделе аспирантуры (каб.402а) Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук» (СПб ФИЦ РАН) по адресу: 199178, Санкт-Петербург, 14-я линия В.О., 39, и на сайте <http://www.spiras.nw.ru/dissovet/>

Автореферат разослан «27» января 2023 г.

Учёный секретарь
диссертационного совета 24.1.206.01,
кандидат технических наук



АБРАМОВ
Максим Викторович

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность. В современном мире всё активнее начинают применяться технологии искусственного интеллекта (ИИ) в совершенно различных областях человеческой деятельности – от анализа больших данных до систем распознавания объектов на фото- и видеоизображениях и автоматических систем контроля и управления, базирующихся на использовании технологий киберфизических систем, промышленного интернета вещей, нейросетевых систем.

Всё это приводит к тому, что становится весьма актуальной **задача** создания систем автоматизации процесса генерации и обучения искусственных нейронных сетей (ИНС) для тех или иных областей. В дальнейшем для обобщённого описания процессов генерации и обучения моделей ИНС будет применяться термин «синтез» моделей ИНС. Всё острее проявляется потребность в системе, подходящей для решения типовых задач из разных областей. При этом, процесс развития и внедрения ИНС в промышленное программное обеспечение (ПО) и повседневную жизнь человека ограничивают как минимум три серьёзных фактора.

Во-первых, существует огромное разнообразие архитектур ИНС, которые, на сегодняшний день, могут применяться для совершенно разных задач, но набор которых при этом слабо систематизирован.

Во-вторых, не существует универсального подхода к генерации и обучению данных архитектур. Разные классы ИНС генерируются и обучаются совершенно разными алгоритмами.

В-третьих, даже уже имеющиеся решения часто реализованы в виде прототипов. Внедрение подобных прототипов будет требовать их глубокой модификации под конкретную задачу, а также написания программных интерфейсов.

Указанные проблемы, а также переход к парадигме Software 2.0, обуславливают **актуальность и новизну** избранной темы диссертации, которая нацелена на автоматизацию процессов генерации, обучения и использования моделей ИНС прямого распространения (ПР) различных архитектур, за счёт создания унифицированного алгоритма синтеза моделей и реализации концепции No-Code разработки. Вторым важным направлением диссертационного исследования является упрощение интеграции моделей ИНС ПР в стороннее ПО, с помощью использования сервис-ориентированной архитектуры (СОА) в процессе генерации программных оболочек для созданных и обученных моделей.

Степень разработанности темы. Подход к автоматизации процессов генерации, обучения и использования моделей ИНС ПР не является новым и уже есть ряд работ на данную тему. Все они сводятся к тому, что автоматизация синтеза моделей машинного обучения (МО) позволит ускорить процесс разработки программных продуктов для решения множества прикладных задач.

Также, на сегодняшний день активно разрабатываются программные комплексы и библиотеки автоматизации обучения ИНС, в том числе и в России. Однако, существующие аналоги зачастую нацелены на решение каких-то узкоспециализированных задач и не рассматривают единообразный и универсальный подход к автоматизации обучения различных архитектур.

К моменту начала работы над диссертацией, научная область, связанная с разработкой алгоритмов автоматизации процессов генерации и обучения моделей ИНС ПР, ещё только зарождалась. Этапы создания и эксплуатации моделей МО, требующие автоматизации, а также основные способы автоматизации этих этапов были сформулированы и опубликованы только в 2014 году международной командой во главе с лабораторией машинного обучения Фрайбургского Университета Альберта-Людвига. Представленный в данной диссертации комплексный подход к автоматизации синтеза моделей ИНС ПР является одним из возможных подходов к решению описанных выше научно-технических задач. На данный момент множество групп разработчиков занимаются исследованием схожих задач. Но многие из них акцентируются на рассмотрении только конкретных задач (к примеру, на свёрточных нейронных сетях) и не используют комплексный подход к автоматизации синтеза моделей ИНС ПР. Другие же научные группы ещё не представили итог своих исследований. Также, схожие работы ведутся научными группами крупных коммерческих компаний. Но данные работы являются частично или полностью закрытыми, в связи с чем можно провести лишь ограниченное сравнение с ними.

Цель исследования заключается в повышении степени автоматизации процесса создания, обучения и использования моделей ИНС прямого распространения различных архитектур. Цель диссертационной работы достигается на основе комплексного решения следующих **научно-технических задач**:

- разработка алгоритма решения задачи автоматизации процессов генерации и обучения моделей ИНС прямого распространения различных архитектур;
- разработка архитектуры и программной системы автоматизации синтеза моделей ИНС прямого распространения с различными архитектурами;
- разработка архитектуры и программной системы автоматической генерации программных оболочек, поддерживающих парадигму сервис-ориентированного подхода и инкапсулирующих алгоритмы работы с созданными моделями ИНС.

Объектом исследования является процесс генерации, обучения и использования созданных моделей ИНС ПР в стороннее ПО.

Предметом исследования является новое программно-математическое обеспечение комплексной автоматизации синтеза моделей ИНС ПР, реализующих различные архитектуры и используемых для решения прикладных задач.

Научная новизна исследования результатов, полученных при решении поставленных задач, состоит в следующем:

1. Разработан алгоритм унифицированного подбора гиперпараметров (структурных параметров) для решения задачи автоматизации процессов генерации и обучения моделей ИНС прямого распространения различных архитектур, отличающийся масштабируемостью по отношению к новым архитектурам ИНС прямого распространения без необходимости модификации самого алгоритма.

2. Разработана архитектура и программная система автоматизации процессов генерации и обучения моделей ИНС прямого распространения с различными

архитектурами, отличающаяся возможностью инкапсуляции методов работы с различными архитектурами ИНС, что позволяет упростить и ускорить разработку моделей за счёт использования перспективной информационной технологии «программирования без программирования» (No-Code разработка).

3. Разработана архитектура и программная система автоматической генерации исполняемых файлов для синтезированных моделей ИНС прямого распространения с различными архитектурами, отличающаяся от существующих конструктивным использованием сервис-ориентированного подхода, а также концепции и технологии No-Code разработки, что позволяет ускорить и упростить интеграцию разработанных моделей в стороннее программное обеспечение.

Теоретическая и практическая значимость диссертации заключается в разработке методов, алгоритмов и их программной реализации, которые определяют основное содержание предложенного в диссертации нового подхода к комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур. В рамках данного подхода удалось создать архитектуру программной системы комплексного автоматизированного синтеза моделей ИНС ПР, которая характеризуется высокой степенью универсальности и унификации и поэтому может использоваться для комплексной автоматизации синтеза моделей различных архитектур ИНС ПР, в том числе не включённых по умолчанию в данное ПО. Также, реализована архитектура программной системы автоматической генерации программных оболочек для синтезированных моделей, которая упрощает и ускоряет внедрение созданных моделей в стороннее ПО, за счёт реализации концепции No-Code разработки и поддержки парадигмы СОА, что подтверждается результатами экспериментальных исследований.

Методология и методы исследования. Для решения поставленных задач в работе используются теория эволюционных алгоритмов, теории модульного программирования и теории глубокого обучения. При компьютерной реализации использовались методы и алгоритмы глубокого обучения, используемые в программных библиотеках Keras и TensorFlow, а также парадигмы No-Code разработки и сервис-ориентированной архитектуры (реализованной с помощью программной библиотеки Flask).

Положения, выносимые на защиту:

1) Алгоритм унифицированного подбора гиперпараметров для решения задачи автоматизации процессов генерации и обучения моделей ИНС прямого распространения различных архитектур.

2) Архитектура и программная система автоматизации процессов генерации и обучения моделей ИНС прямого распространения различных архитектур.

3) Архитектура и программная система автоматической генерации исполняемых файлов для синтезированных моделей ИНС прямого распространения различных архитектур.

Соответствие диссертации специальности. Представленные результаты соответствуют специальности 2.3.5 – Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей.

Степень достоверности результатов диссертации обеспечивается анализом состояния современных исследований по тематикам МО и автоматизированного обучения систем ИИ, подтверждается согласованностью полученных результатов, успешной апробацией созданного ПО и последующим внедрением результатов его работы, а также 11 выступлениями на международных и российских научных конференциях и публикацией итогов исследований в ведущих рецензируемых изданиях.

Апробация и реализация результатов. Модели ИНС, созданные в результате использования разработанного ПО, были использованы при реализации информационной системы (ИС) ПРОСТОР. Исследования, отражённые в диссертации, проводились в рамках 4 научно-исследовательских работ: 1) грант РФФИ №19-37-90112 «Разработка методов, технологии и программного комплекса автоматизированной генерации и обучения искусственных нейронных сетей на основе сервис-ориентированной архитектуры»; 2) грант РФФИ №19-08-00989 «Разработка и исследование научных основ теории многокритериального оценивания, анализа и управления качеством моделей и полимодельных комплексов, описывающих сложные технические объекты»; 3) грант РФФИ №17-08-00797 «Разработка и исследование методологических основ и технологии комплексного моделирования процессов функционирования системы проактивного управления сложными техническими объектами»; 4) грант РФФИ №16-08-00510 «Разработка и исследование методологии построения и создание прототипа информационной автоматизированной системы прогнозирования состояния растительного покрова Крайнего Севера на основе интегрированной обработки мульти- и гиперспектральных наземно-аэрокосмических данных, а также климатической информации».

Основные результаты работы докладывались и обсуждались на 11 научно-практических конференциях: «32nd European Conference on Modelling and Simulation» (ECMS 2018), г. Вильгельмсхафен, Германия, 2018 г.; «20th International Conference on Harbor, Maritime & Multimodal Logistics Modelling and Simulation» (HMS2018), г. Будапешт, Венгрия, 2018 г.; «Глобальные климатические изменения: региональные эффекты, модели, прогнозы», г. Воронеж, 2019 г.; «33rd European Conference on Modelling and Simulation» (ECMS 2019), г. Вильгельмсхафен, Германия, 2019 г.; «13th IEEE International Conference “Application of Information and Communication Technologies”» (AICT2019), г. Баку, Азербайджан, 2019 г.; «2nd Euro-Mediterranean Conference for Environmental Integration» (EMCEI-2), г. Сус, Тунис, 2019 г.; «9th IFAC Conference on Manufacturing Modelling, Management and Control» (MIM 2019), г. Берлин, Германия, 2019 г.; «Модели и методы исследования информационных систем на транспорте» (ММРИСТ-2020), г. Санкт-Петербург, 2020 г.; «18 Национальная Конференция по Искусственному Интеллекту с Международным Участием» (КИИ-2020), г. Москва, 2020 г.; «Математическое моделирование в экологии» (ЭкоМатМод-2021), г. Пушино, 2021 г.; «Pattern Recognition and Information Processing» (PRIP'2021), г. Минск, Беларусь.

Публикации. По теме диссертации опубликовано 25 печатных работ, включая 3 публикации в журналах из перечня рецензируемых научных изданий, в которых

должны быть опубликованы основные научные результаты диссертаций на соискание учёной степени кандидата наук, 10 публикаций в изданиях, индексируемых в WoS/Scopus, 2 свидетельства о регистрации ПрЭВМ в Роспатент. Полный перечень публикаций соискателя по теме исследования представлен в Приложении А диссертационной работы.

Личный вклад автора. Основные научные положения, теоретические выводы и практические решения, результаты тестирования сформулированы и изложены автором самостоятельно.

Структура и объём работы. Диссертация объемом 144 машинописные страницы содержит введение, четыре главы и заключение, список литературы (148 наименований), 1 таблицу, 32 рисунка, приложение со списком публикаций соискателя.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении дана общая характеристика работы, обоснована актуальность проводимых исследований и разработок, определены цели работы, основные задачи исследования, отражена научная новизна и практическая значимость работы, сформулированы результаты, выносимые на защиту.

В первой главе проведён анализ современного состояния исследований проблем практического применения ИНС, который показал, что потребность в программных решениях на базе МО в целом и ГО в частности растёт быстрее количества специалистов, способных решать соответствующие прикладные задачи, используя методы и технологии МО. Всё это приводит к необходимости комплексной автоматизации процессов генерации и обучения моделей МО и ГО, а также создания инструментов, упрощающих последующую интеграцию созданных решений в стороннее ПО.

Существующие же подходы (методологии и методики) и соответствующие инструментальные средства автоматизации процессов генерации и обучения моделей ИНС хоть и позволяют решать точно отдельные из перечисленных проблем, однако они не обеспечивают реализацию комплексного подхода. Их использование все ещё требует наличия у специалистов знаний о специфике генерации и обучения моделей МО и ГО, существует привязка к конкретным платформам или же отсутствует возможность масштабирования, из-за специфичной реализации. Так что, несмотря на имеющиеся достижения в решении отдельных задач, нельзя сказать, что на сегодняшний день проблема автоматизации процессов генерации, обучения и использования моделей ИНС решена в полной мере.

За основу постановки задачи повышения степени автоматизации процессов создания, обучения и использования моделей ИНС прямого распространения различных архитектур, был взят ГОСТ 23004-78 «Механизация и автоматизация технологических процессов в машиностроении и приборостроении». С учетом содержания данного документа формальная постановка целевой задачи может быть представлена следующим образом:

$$\text{card } A(a, L^a(p, g^a)) \rightarrow \max_{A \subseteq \dot{P}(O)}, \quad (1)$$

где A – множество автоматизированных первичных частей (например, этапов технологических процессов) процесса создания, обучения и использования моделей ИНС прямого распространения различных архитектур;

a – индекс (номер) архитектуры ИНС ПР, используемой для решения поставленной пользователем задачи;

L^a – функция (алгоритм) автоматического подбора гиперпараметров ИНС ПР архитектуры a ;

p – обучающая выборка (включающая в себя данные для валидации моделей), предоставленная пользователем и используемая в процессах генерации и обучения моделей ИНС ПР различных архитектур;

g^a – набор гиперпараметров модели ИНС ПР архитектуры с номером a , варьируемых в процессе обучения;

O – множество всех первичных частей процесса создания, обучения и использования моделей ИНС прямого распространения различных архитектур;

$\dot{P}(O)$ – булеан множества всех первичных частей процесса создания, обучения и использования моделей ИНС прямого распространения различных архитектур.

Тогда, определение степени автоматизации процессов создания, обучения и использования моделей ИНС прямого распространения различных архитектур можно вычислить по формуле:

$$D = \frac{\text{card } A(a, L^a(p, g^a))}{\text{card } O(a, L^a(p, g^a))}, \quad (2)$$

где D – показатель степени автоматизации первичных частей технологического процесса создания, обучения и использования моделей ИНС прямого распространения различных архитектур.

Для достижения поставленной выше задачи необходимо отдельно решить подзадачу автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур. Формальная постановка этой задачи имеет следующий вид:

$$Q^a(N_i^a(p, L^a(p, g^a)), q) \rightarrow \max_{i \in [1, n]}, \quad (3)$$

при выполнении условия:

$$P_i(N_i^a) \geq P_{\text{зад}},$$

где p – обучающая выборка (включающая в себя данные для валидации моделей), предоставленная пользователем и используемая в процессах генерации и обучения моделей ИНС ПР различных архитектур;

q – тестовая выборка, предоставленная пользователем как входные данные для разрабатываемой программной системы автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур;

a – индекс (порядковый номер) архитектуры ИНС ПР, используемой для решения поставленной пользователем задачи;

g^a – набор гиперпараметров модели ИНС ПР архитектуры с номером a , варьируемых в процессе обучения;

L^a – функция (алгоритм) автоматического подбора гиперпараметров ИНС ПР архитектуры a ;

N_i^a – i -ый экземпляр модели ИНС ПР архитектуры a , который был обучен на выборке p и с гиперпараметрами, полученными в процессе работы

соответствующей функции (алгоритма) автоматического подбора гиперпараметров архитектуры ИНС (L^a);

i – индекс экземпляра модели ИНС ПР;

n – заданное количество экземпляров генерируемых и обучаемых моделей ИНС ПР, среди которых производится поиск модели;

Q^a – функция расчёта качества (в том числе, и точности) работы i -ого экземпляра модели ИНС ПР архитектуры a (N_i^a) на тестовой выборке q ;

P_i – функция расчёта точности обученной модели;

$P_{\text{зад}}$ – заданное пользователем граничное значение требуемой точности обученной модели.

Перечислены и обоснованы требования к алгоритмам автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур. Сформулировано требование к модульности создаваемого ПО, что позволило решить проблему унифицированного подхода к различным архитектурам ИНС, с одной стороны, и проблему масштабируемости, с другой стороны. Обосновано применение парадигмы СОА при генерации программных оболочек для созданных моделей, что позволило на конструктивном уровне решить вопрос их интеграции в стороннее ПО.

Во второй главе предложены архитектуры ИНС ПР, использование которых оправдано в рамках разрабатываемого ПО. Для каждой архитектуры, во-первых, обоснована причина её использования в разрабатываемом ПО и указана область применения, во-вторых, описаны особенности её генерации, обучения и функционирования, в-третьих, представлены методы её интеграции в разрабатываемый ПО, с учётом описанных особенностей. Представленные в данной главе архитектуры предоставляются пользователю на выбор, при потребности в решении соответствующих задач.

Многослойный перцептрон Румельхарта (МПР) – несмотря на то, что архитектура была описана уже несколько десятков лет назад, она до сих пор находит применение при решении многих задач. Основными достоинствами данной архитектуры ИНС являются её простота и универсальность.

При автоматизации обучения моделей данной архитектуры ИНС ПР предполагается варьировать следующие гиперпараметры: количество скрытых слоёв, количество нейронов в скрытых слоях, количество эпох обучения, функции активации скрытых слоёв, алгоритм оптимизации нейронной сети, скорость обучения, количество шагов в эпохе обучения. Варьирование более специфических гиперпараметров сети (к примеру коэффициент функции активации Leaky ReLU) не предполагается, поскольку это является избыточным для поставленной задачи.

На текущем этапе разработки выбор наиболее точной модели будет осуществляться на основе значения среднеквадратичной ошибки. Но в дальнейшем предполагается более широкий выбор методов оценки точности модели, в т.ч. пользовательские.

Свёрточная нейронная сеть (СНС) является более узкоспециализированным инструментом, подходящим прежде всего для работы с изображениями и иными данными, которые возможно представить в матричном виде. Одной из первых и наиболее известной является модель, разработанная Яном Лекуном. В основу

логики функционирования данного класса моделей были положены исследования зрительной коры головного мозга животных и созданные на их основе алгоритмы распознавания. Данная архитектура предназначена для автоматического и адаптивного изучения многомерных зависимостей в массиве входных данных. Причём СНС учитывает как низкоуровневые зависимости, так и высокоуровневые.

Для автоматизации обучения моделей данной архитектуры предполагается варьировать: количество свёрточных слоёв, количество полносвязных слоёв, количество нейронов в свёрточных слоях, количество нейронов в полносвязных слоях, включение/выключение дропаута, количество эпох обучения, скорость обучения.

На текущем этапе разработки выбор наиболее точной модели будет осуществляться на основании значения метрики F-меры. Но в дальнейшем предполагается более широкий набор методов оценки точности модели, в т.ч. пользовательские.

СНС сегментации и семантического анализа изображения является наиболее узкоспециализированной из имеющихся в системе архитектур ИНС. Mask Regions with Convolution Neural Networks (MRCNN) является подвидом классической СНС и, за счёт усложнения архитектуры, она успешно справляется с задачами семантической и объектной сегментации изображений.

Процессы генерации и обучения моделей архитектуры MRCNN концептуально не отличаются от процессов генерации и обучения обычной СНС. Однако, поскольку архитектура MRCNN является уже фиксированной и оптимизированной, при автоматизированном обучении не предполагается модификация самой архитектуры. Поэтому для варьирования остаются только следующие гиперпараметры – количество эпох обучения, количество шагов в эпохе, скорость обучения.

Тем не менее, из-за сложности обучения данной архитектуры, она имеет уникальную для неё опцию – возможность предобучения. Такой подход распространён при обучении масштабных моделей ИНС и носит название «трансферное обучение». В случае выбора данной опции обучение сети в разработанном ПО будет производиться не с нуля. Вместо этого за основу будет взята модель MRCNN уже обученная на массиве данных Microsoft COCO Dataset (Microsoft Common Objects in Context). Использование данного массива для базового обучения MRCNN позволяет задать для неё все основные концепции различных классов объектов.

На текущем этапе разработки, из-за специфики архитектуры СНС, выбор наиболее точной модели будет осуществляться на основании пользовательского алгоритма оценки точности модели. Поскольку универсальные методы оценки для данной модели отсутствуют.

Разработан **алгоритм унифицированного подбора гиперпараметров (УПГ)** для автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур, который предназначен для использования в разрабатываемом ПО в качестве основного алгоритма генерации и обучения моделей ИНС. Инкапсуляция логики функционирования моделей ИНС ПР различных архитектур и модульная структура ПО, предложенные в диссертационной работе, позволяют единообразно

обращаться к различным архитектурам и обучать их с помощью одного программного модуля.

Алгоритм УПГ является модификацией генетического алгоритма, но нацелен прежде всего на унифицированный подход к подбору гиперпараметров различных архитектур ИНС ПР, и имеет следующий вид:

Шаг 1. В первой родительской популяции генерируется фиксированное число (M) моделей ИНС фиксированной архитектуры (выбранной автоматически на основе представленных пользователем входных данных), со случайно заданными гиперпараметрами.

Шаг 2. Генерируется N_d дочерних моделей ИНС фиксированной архитектуры, гиперпараметры которых выбираются равновероятным случайным образом из гиперпараметров двух случайно подобранных родительских моделей ИНС (для двух родительских моделей – одна дочерняя), а также генерируется N_r моделей ИНС фиксированной архитектуры, гиперпараметры которых задаются полностью случайно, с учётом заданных диапазонов значений для этих гиперпараметров. Оператор мутации (инвертирования генов) для дочерних моделей ИНС не применяется из-за специфического формата «генотипа».

Шаг 3. Далее осуществляется селекция моделей ИНС по методу рулетки:

$$p_i = \frac{f_i}{\sum_{j=1}^N f_j}, \quad (4)$$

где p_i – вероятность выбора i -ой модели ИНС;

f_i – значение функции приспособленности для i -ой модели ИНС;

N – количество моделей ИНС в популяции на текущем шаге (формула 5).

$$N = M + N_d + N_r, \quad (5)$$

Метод рулетки выбран из-за его универсальности и простоты программной реализации, поскольку алгоритм УПГ будет использоваться для разных классов задач. Использование узкоспециализированных алгоритмов хоть и повысило бы скорость работы для одних классов задач, но неизбежно понизило бы скорость работы для других. В основу функции приспособленности положен расчёт погрешности значения целевого параметра, рассчитанного с помощью моделей ИНС, относительно реального значения тестовой выборки:

$$f_i = \frac{1}{\sqrt{\frac{\sum_{j=1}^X (\varepsilon_{ij} - \omega_j)^2}{X}}}, \quad (6)$$

где ε_{ij} – выходное значение целевого параметра спрогнозированное i -ой моделью ИНС в ответ на j -ый входной тестовый вектор;

ω_j – реальное значение тестовой выборки в ответ на j -ый входной тестовый вектор;

X – количество тестовых векторов.

Результат, рассчитанный по данной формуле, является значением «уровня приспособленности», которое обратно пропорционально среднеквадратической ошибки i -ой модели ИНС на тестовой выборке. В результате селекции, в текущее поколение, из N моделей ИНС, отбирается M моделей, с наибольшим значением p_i (вероятности выбора i -ой модели ИНС).

Шаг 4. Для всех моделей ИНС, вычисляется (по специфической для каждой архитектуры формуле) ошибка рассчитанного ими значения целевого параметра

относительно реального значения тестовой выборки. Если хотя бы одна модель ИНС показывает ошибку ниже заданного значения, цикл прерывается. Модель ИНС с наименьшей ошибкой принимается в качестве «победившей». В противоположном случае, происходит возврат к шагу 2. При этом, популяция моделей каждой итерации отдельно запоминается. Если популяция моделей текущей итерации полностью совпадает с предыдущей популяцией, это означает, что за всю итерацию не было найдено конфигурации модели ИНС с большей точностью и осуществляется безусловный переход к шагу 5.

Шаг 5. Если не найдена модель ИНС с ошибкой меньше заданного значения, цикл запускается с 1 шага с новой родительской популяцией моделей, для которой задаются новые случайные значения параметров. Если за I итераций решение не найдено, задача признаётся нерешаемой, при заданных настройках, и осуществляется выход из алгоритма.

Ключевые отличия УПГ от исходного генетического алгоритма следующие:

- «Генотип» представлен в виде в виде расширяемого массива, элементы которого могут быть как целочисленными, так и вещественными. Это обусловлено тем, что, во-первых, гиперпараметры ИНС могут быть как целочисленными, так и вещественными, а во-вторых, в разработанное ПО могут добавляться новые архитектуры ИНС ПР и было бы сложно (в программной реализации) избежать модификации созданного алгоритма при ином формате «генотипа».

- Оператор мутации (инвертирования генов) для дочерних моделей ИНС не применяется из-за специфического формата «генотипа».

- Кроссовер (оператор скрещивания) является случайным и многоточечным (при каждом скрещивании ген разбивается на количество частей равное количеству элементов массива). Подобная модификация является специфичной для поставленной задачи, поскольку заранее не известно, какие гиперпараметры и каким образом нужно изменять, чтобы повысить точность модели ИНС конкретной архитектуры.

Применение описанного алгоритма УПГ к задаче комплексной автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур позволит решить две ключевые задачи. Во-первых, данный алгоритм позволит единообразно обучать различные модели ИНС ПР, поскольку пул гиперпараметров не фиксирован и может быть адаптирован под разные архитектуры без необходимости модификации самого алгоритма. Так как вся логика работы с конкретными архитектурами ИНС ПР будет инкапсулирована, алгоритм в полностью автоматическом режиме сможет переключаться между различными архитектурами в процессе своей работы. Во-вторых, подход с рекомбинацией гиперпараметров позволит ускорить сам процесс синтеза моделей. Поскольку гиперпараметры новых моделей ИНС будут комбинироваться с гиперпараметрами уже существующих моделей, это позволит быстрее получать модели, обеспечивающие решение поставленной задачи с заданной точностью.

Структурная схема алгоритма УПГ представлена на рисунке 1.

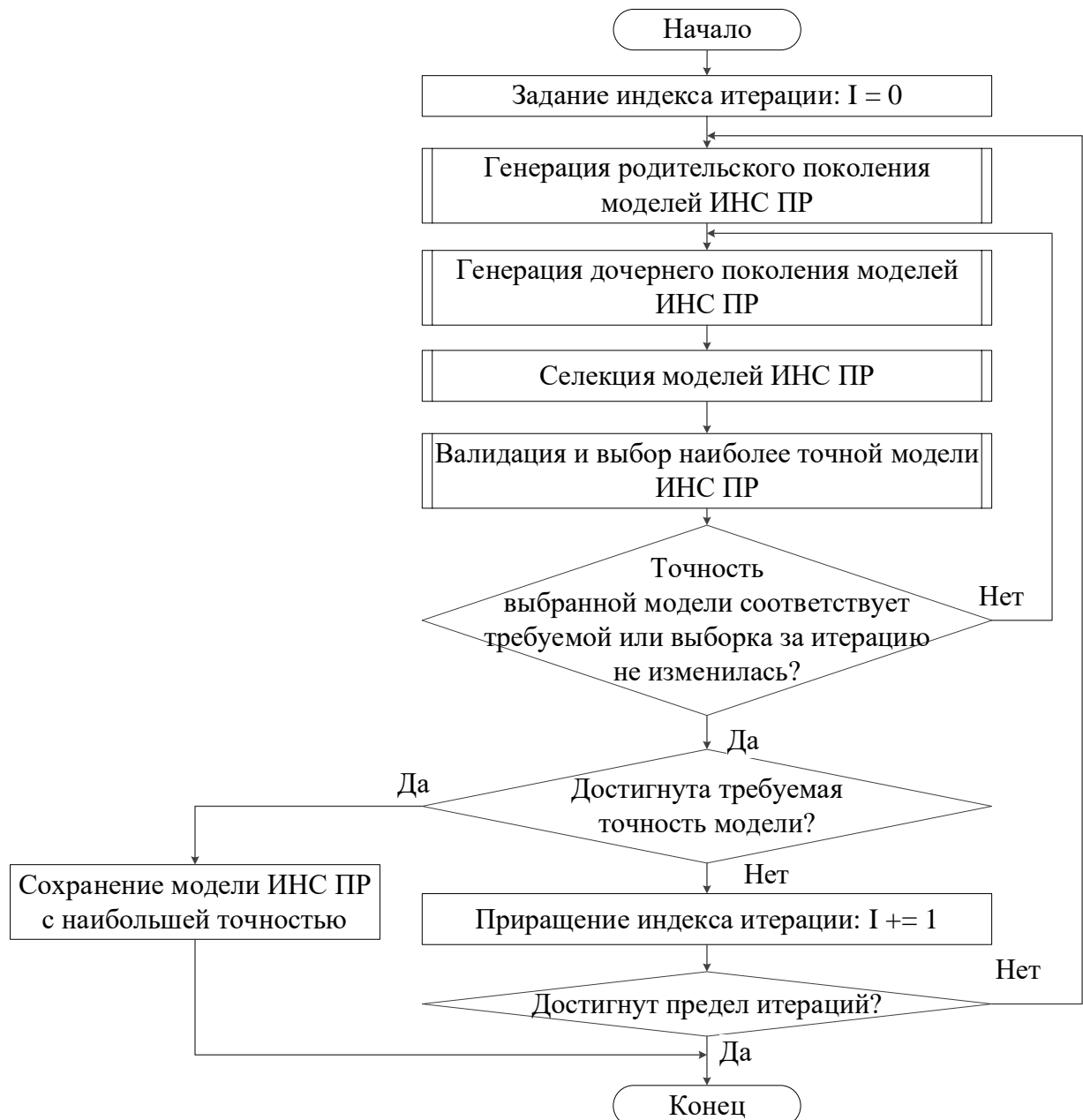


Рисунок 1 – Структурная схема алгоритма УПГ для автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур

В третьей главе разработаны и обоснованы состав, структура и технологии использования созданного ПО комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур. При этом разработана и реализована как общая архитектура спроектированного ПО, так и составляющие его модули. В состав ПО комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур вошло одиннадцать программных модулей, реализующих на конструктивном уровне существующие парадигмы СОА и модульной структуры (рисунок 2). Разработанные программные модули использует шаблон проектирования «Фасад», который позволяет свободно масштабировать и расширять созданное ПО, в том числе за счёт добавления новых архитектур ИНС ПР.



Рисунок 2 – Детальная схема ПО комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур

Помимо разбиения на модули отдельных архитектур ИНС ПР, ПО также условно делится на внешнюю и внутреннюю части. Внешняя часть отвечает за взаимодействие с пользователем и форматирование входных и выходных данных. Внешняя часть ПО подразделяется на следующие модули:

- модуль предобработки входных данных;
- модуль генерации исполняемого файла сгенерированной модели ИНС ПР;
- модуль генерации REST и SOAP оболочек для сгенерированной модели ИНС ПР;
- модуль формирования пакета программ для сгенерированной модели ИНС ПР;
- рекомендательный модуль (отвечает за предоставление рекомендаций пользователю в случае невозможности создания модели ИНС ПР для заданных входных данных);
- пользовательский интерфейс.

Внутренняя часть ПО ориентирована исключительно на генерацию и обучение моделей ИНС ПР на основе входных данных, предоставленных пользователем. Внутренняя часть ПО подразделяется на следующие модули:

- модуль инициализации (отвечает за инициализацию исходных условий решаемой задачи, в т.ч. родительского поколения моделей ИНС ПР);

- модуль генерации итеративного поколения моделей;
- модуль валидации и отбора моделей;
- модуль итоговой валидации и верификации модели;
- модули описывающие архитектуры ИНС ПР (в т.ч. все функции необходимые для взаимодействия с моделью конкретной архитектуры и варьирования её параметров).

Принципиальная отличительная особенность разработанного ПО заключается в комплексной автоматизации всех этапов генерации и обучения моделей ИНС ПР различных архитектур, а также этапов генерации программных и сервисных оболочек для созданных моделей. Созданные программные модули в автоматизированном режиме, согласно указанным пользователем настройкам, обеспечивают реализацию полного цикла проектирования и генерации моделей ИНС: от базовой нормализации обучающих данных и выбора гиперпараметров модели до настройки параметров процесса обучения. В результате работы ПО формируется пакет кроссплатформенных программ, который позволяет свободно использовать созданную модель ИНС для решения практических задач напрямую или интегрировать её в существующее ПО и обращаться к ней через программный интерфейс. Пользовательский интерфейс ПО предоставляет необходимый минимум инструментов для автоматизации процесса создания моделей ИНС ПР различных архитектур. Таким образом, вся сложность процессов генерации и обучения моделей ИНС скрыта от пользователя за счёт автоматизации данных процессов. За счёт этого в созданном ПО реализуется концепция No-Code разработки, благодаря которой пользоваться им смогут не только специалисты с высоким уровнем знаний в области МО, но и специалисты из различных прикладных областей, не имеющие компетенций по вопросам использования МО и ГО, а также другие пользователи, заинтересованные в создании моделей ИНС ПР.

Были введены механизмы автоматизации процессов интеграции создаваемых моделей ИНС в стороннее ПО на основе использования парадигмы SOA. Данный подход не привязан к конкретному ПО и сетевому окружению, поэтому может быть реализован с помощью различных технологий. При этом к созданным моделям ИНС ПР можно будет обращаться единообразно из различных программных сред.

Согласно ГОСТ Р 57412-2017 «Компьютерные модели в процессах разработки, производства и эксплуатации изделий» этапы разработки компьютерной модели ИНС ПР можно сформулировать следующим образом:

- 1) содержательная постановка задачи на концептуальном уровне;
- 2) построение модели ИНС ПР;
- 3) выбор метода решения с учётом знаний и предпочтений пользователя и разработчика;
- 4) программная реализация модели ИНС ПР;
- 5) применение полученной модели ИНС ПР при решении прикладных задач;
- 6) контроль и анализ полученных результатов.

Поскольку большинство существующих средств AutoML являются системами автоматизации работы специалистов по машинному обучению, они автоматизируют только пункты 2 и 4. Поэтому на практике показатель степени

автоматизации первичных частей процесса создания, обучения и использования моделей ИНС ПР различных архитектур у таких систем оказывается $D=0.33$.

С другой стороны, в разработанном ПО автоматизируются в том числе этапы 3, 5 и 6. Поэтому для ряда прикладных задач показатель степени автоматизации первичных частей процесса создания, обучения и использования моделей ИНС ПР различных архитектур разработанного ПО может достигать $D=0.83$.

В разработанное ПО были интегрированы сторонние программные библиотеки и программно-аппаратные ускорители, позволяющие реализовать выбранные во 2 главе алгоритмы и методы, которые требуют больших вычислительных ресурсов. Без использования предложенных программных библиотек и программно-аппаратных ускорителей решение практических задач могло бы занимать месяцы, что полностью лишало бы смысла использование представленного ПО. Всё это повышает универсальность созданного ПО и существенно упрощает процесс интеграции синтезированных моделей ИНС ПР.

Новым практическим результатом, полученным в данной главе, стало конкретное доказательство реализуемости концепции комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур, предложенной в предыдущих главах диссертации. Было продемонстрировано, что за счёт использования модульной архитектуры и парадигмы СОА можно создать ПО, которое бы единообразно осуществляло взаимодействие с принципиально различными архитектурами ИНС ПР.

В четвёртой главе была разработана методика использования разработанного ПО комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур, с помощью которой были решены три задачи из принципиально разных областей. С использованием представленной методики в автоматизированном режиме были синтезированы различные модели ИНС ПР, каждая из которых отличалась как процессами генерации и обучения, так и процессом их интеграции. При этом все эти модели генерировались и обучались с помощью одного и того же ПО, на основе разработанного алгоритма УПГ для автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур. Результаты экспериментальных исследований подтвердили повышение оперативности создания моделей ИНС ПР, достигаемые при использовании разработанного ПО.

При создании моделей прогнозирования уровня воды, за счёт использования разработанного ПО, было снижено общее время их разработки с нескольких недель до нескольких суток (по результатам экспериментальной оценки). Практически до нуля было снижено время необходимое для интеграции созданных моделей в ИС «ПРОСТОР» (рисунок 3), поскольку с помощью разработанного ПО сразу были сгенерированы сервисные оболочки необходимые для подключения к общей сервисной шине ИС. Показатель степени автоматизации первичных частей процесса создания, обучения и использования моделей ИНС ПР различных архитектур при решении данной задачи составил $D=0.83$.

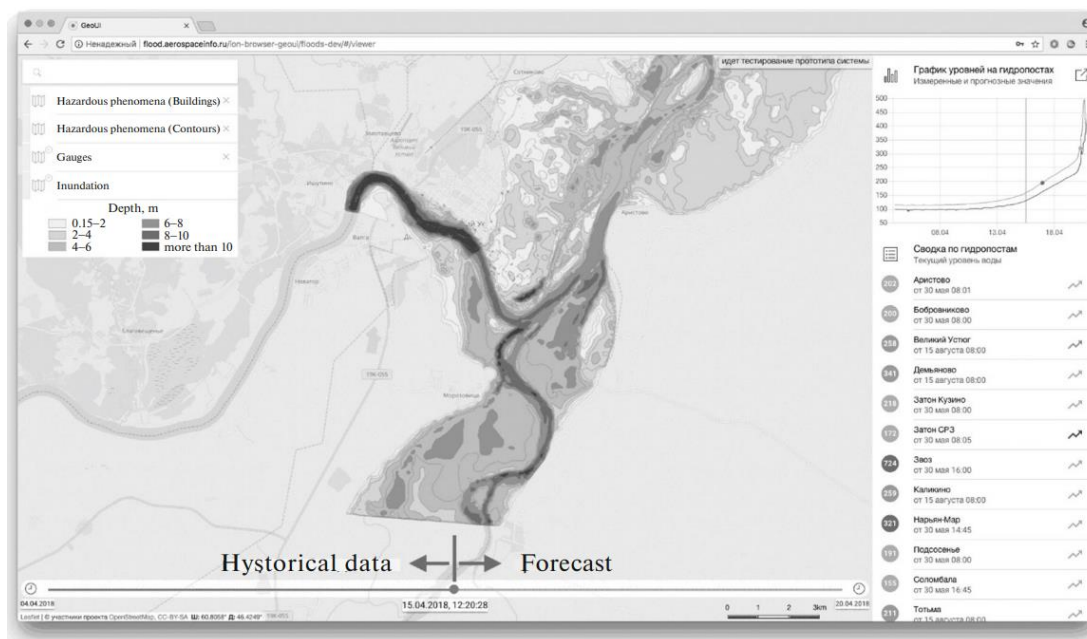


Рисунок 3 – Пример интерфейса ИС «ПРОСТОР» с результатами краткосрочного прогнозирования уровня воды для конкретного гидропоста (правый верхний угол интерфейса)

При решении задачи прогнозирования динамики изменения фитомассы растительных сообществ тундры было сокращено время необходимое на прототипирование модели. Был осуществлён автоматизированный перебор гиперпараметров моделей ИНС ПР, благодаря чему рабочий прототип был получен за сутки, вместо нескольких рабочих дней (по результатам экспериментальной оценки). Показатель степени автоматизации первичных частей процесса создания, обучения и использования моделей ИНС ПР различных архитектур при решении данной задачи составил $D=0.5$.

При синтезе моделей ИНС ПР распознавания и подсчёта числа северных оленей по аэрофотоснимкам (рисунок 4) разработанное ПО позволило больше всего сэкономить время на указанный синтез. Поскольку модели СНС в процессе обучения являются одними из наиболее требовательных к затратам вычислительных ресурсов, предложенный подход к комплексной автоматизации процесса этого обучения позволил созданному ПО надёжно функционировать круглые сутки без остановки. При этом время на создание модели было сокращено с нескольких месяцев до нескольких суток (по результатам экспериментальной оценки). Показатель степени автоматизации первичных частей процесса создания, обучения и использования моделей ИНС ПР различных архитектур при решении данной задачи составил $D=0.67$.

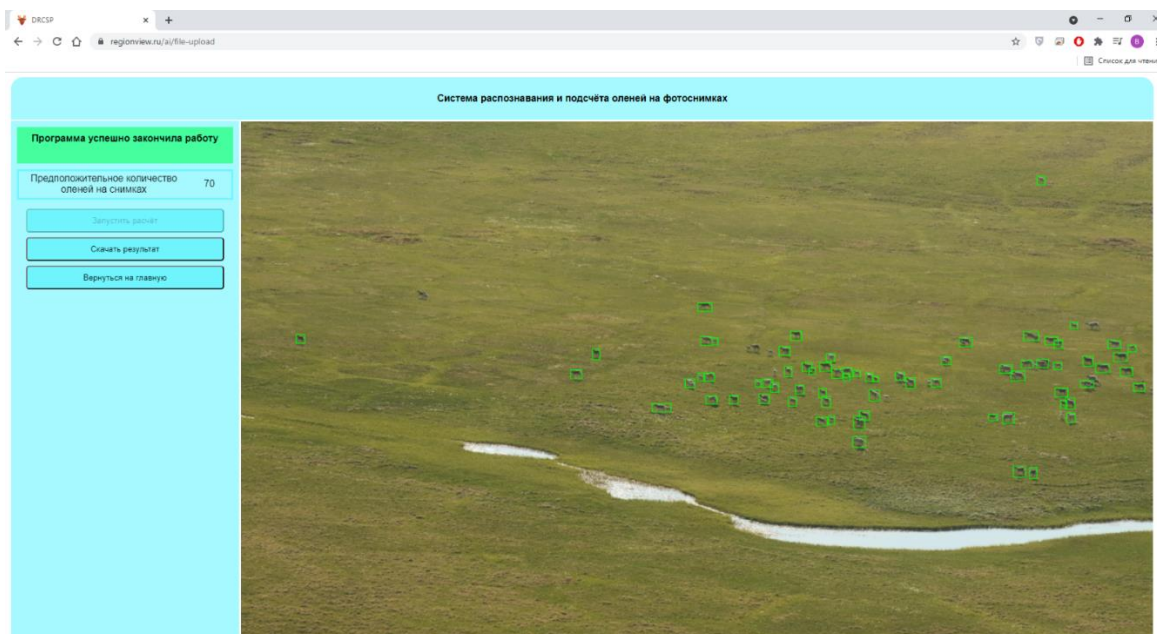


Рисунок 4 – Пример веб-интерфейса разработанного ПО автоматического распознавания и подсчёта количества северных оленей

В главе наглядно продемонстрировано конструктивное использование СОА при генерации программных оболочек для созданных моделей ИНС ПР. Результаты интеграции в ИС «ПРОСТОР» наиболее ярко показали необходимость и оправданность выбранного подхода для упрощения процесса интеграции созданных моделей ИНС ПР в стороннее ПО

ЗАКЛЮЧЕНИЕ

В представленной диссертации сформулирована и решена новая научно-техническая задача комплексной автоматизации процессов генерации, обучения и использования моделей ИНС ПР различных архитектур. Решённая задача имеет важное значение для совершенствования методов и алгоритмов автоматизированного машинного обучения, используемых для упрощения и ускорения процесса разработки моделей ИНС ПР, что в свою очередь приводит к удешевлению и ускорению разработки программных комплексов на базе моделей ИНС, решающих прикладные задачи в различных областях человеческой деятельности. Результаты апробации разработанного ПО подтверждают достижение цели диссертационной работы – повышения степени автоматизации процесса создания, обучения и использования моделей ИНС ПР различных архитектур за счёт автоматизации их разработки, базирующейся на целенаправленном выборе архитектур и гиперпараметров создаваемых моделей ИНС ПР с использованием алгоритма УПГ, а также автоматической генерации программных оболочек для синтезированных моделей ИНС ПР.

Основные научные результаты, составляющие **итоги** исследования:

1. Разработан алгоритм УПГ для автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур. Предложенный алгоритм обеспечивает унифицированный подход к генерации и обучению моделей ИНС ПР различных архитектур, который отличается от существующих подходов к решению рассматриваемой задачи синтеза тем, что при его реализации отсутствует

необходимость внесения модификаций в сам алгоритм при работе с каждой конкретной архитектурой ИНС ПР. Эта возможность достигнута за счёт внесения в исходный генетический алгоритм модификаций, учитывающих обобщённые особенности подхода обучения с учителем к ИНС ПР.

2. Разработана архитектура и программная система автоматизации процессов генерации и обучения моделей ИНС ПР различных архитектур, отличающаяся модульной расширяемой структурой, что позволяет использовать её для решения широкого спектра практических задач. Программная система в составе ПО была апробирована на ряде прикладных задач и показала свою вычислительную эффективность и удобство при её использовании непрограммирующими пользователями. Благодаря реализации концепции No Code в разработанном ПО, конечный пользователь способен создавать полностью работоспособные программные реализации моделей ИНС ПР различных архитектур, без необходимости самостоятельно писать программный код.

3. Разработана архитектура и программная система автоматической генерации программных оболочек для созданных моделей, с учётом парадигмы SOA. Данная программная система позволяет ускорить и упростить интеграцию созданных моделей в стороннее ПО. Благодаря реализации концепции No Code в разработанной программной системе, конечный пользователь способен создавать интегрируемые программные модули для созданных моделей ИНС ПР реализующих различные архитектуры, без необходимости самостоятельно писать программный код.

Перспектив дальнейшей разработки темы включают вопрос развития пользовательского программного интерфейса, который бы позволил работать с созданным ПО неспециалистам в области ГО (в данный момент реализован лишь рабочий прототип интерфейса). Кроме этого, требуется ввести разработанное ПО в ограниченную эксплуатацию, для оценки вычислительных ресурсов, необходимых для его штатной работы, а также для проведения стресс-тестов и тестов производительности. После проведения всех необходимых тестов и оценки разработанного ПО, планируется его дальнейшая полноценная эксплуатация, с последующим расширением набора архитектур ИНС ПР, комплексную автоматизацию процессов генерации, обучения и использования которых он сможет обеспечить. Ограничения разработанного ПО, на сегодняшний день, связаны с небольшим количеством реализованных в ней архитектур ИНС ПР. Несмотря на то, что на данный момент разработанное ПО обеспечивает решение наиболее востребованных на практике прикладных задач, при дальнейшей разработке необходимо будет увеличить количество доступных архитектур ИНС ПР.

Полученные результаты соответствуют паспорту специальности 2.3.5 «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей» по техническим наукам.

Разработанное алгоритмическое и программное обеспечение реализовано в ряде организаций: Государственный природный биосферный заповедник «Таймырский», «Санкт-Петербургский государственный технологический институт (технический университет)» СПбГТИ (ТУ), в СПИИРАН и СЗЦПО входящих в СПб ФИЦ РАН.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в журналах из перечня рецензируемых научных изданий, в которых должны быть опубликованы основные научные результаты диссертаций на соискание учёной степени кандидата наук, на соискание учёной степени доктора наук:

1. Соболевский В.А. Автоматизированная система генерации, обучения и использования искусственных нейронных сетей // Информатизация и связь. 2019. №3. С. 100-107. DOI:10.34219.2078-8320-2019-10-3-100-107.
2. Соболевский В.А. Сервис-ориентированный подход к разработке систем на базе свёрточных нейронных сетей // Информатизация и связь. 2020. №5. С.34–40. DOI: 10.34219/2078-8320-2020-11-5-34-40.
3. Михайлов В.В., Соболевский В. А., Колпашиков Л. А., Соловьев Н. В., Якушев Г. К. Методологические подходы и алгоритмы распознавания и подсчета животных на аэрофотоснимках // Информационно-управляющие системы. 2021. №5 (114). С. 20-32. DOI: 10.31799/1684-8853-2021-5-20-32.

В изданиях, включенных в базу данных Scopus:

1. Mikhailov, V.V., Sobolevskii, V.A., Kolpaschikov, L.A. Mask R-CNN-Based System for Automated Reindeer Recognition and Counting from Aerial Photographs // Communications in Computer and Information Science. 1562. 2022. Стр. 137–151. DOI: 10.1007/978-3-030-98883-8_10.
2. Mikhailov V.V., Spesivtsev A.V., Sobolevsky V.A., Kartashev N.K., Spesivtsev V.A., Lavrinenko I.A., Lavrinenko O.V. Multimodel evaluation of phytomass dynamics of tundra plant communities based on satellite images // Izvestiya, Atmospheric and Oceanic Physics. 57(9). 2021. Стр. 1198-1210. DOI: 10.1134/S0001433821090553.
3. Mikhailov V., Ponomarenko M., Sobolevsky V. Simulation of phytomass dynamics of plant communities based on artificial neural networks and NDVI // Recent Advances in Environmental Science from the Euro-Mediterranean and Surrounding Regions (2nd Edition). EMCEI 2019. Environmental Science and Engineering. 2021. Стр. 1335-1339. DOI: 10.1007/978-3-030-51210-1_211.
4. Sobolevskii V.A. The system of convolution neural networks automated training // CEUR Workshop Proceedings. 2803. 2020. Стр. 100-106. DOI 10.24412/1613-0073-2803-100-106.
5. Gnidenko, A., Sobolevsky, V., Potriasaev, S., Sokolov, B. Methodology and integrated modeling technologies for synthesis of cyber-physical production systems modernization programs and plans // IFAC-PapersOnLine. 52(13). 2019. Стр. 642–647. DOI: 10.1016/j.ifacol.2019.11.305.
6. Rostova, E.N., Rostov, N.V., Sobolevsky, V.A. Synthesis and simulation of biotechnical position-force control system of a robot manipulator with reconfigurable

structure // IFAC-PapersOnLine. 52(13). 2019. Стр. 1097–1101. DOI: 10.1016/j.ifacol.2019.11.342.

7. Zelentsov V.A., Alabyan A.M., Krylenko I.N., Pimanov I.Yu., Ponomarenko M.R., Potryasaev S.A., Semenov A.E., Sobolevskii V.A., Sokolov B. V., Yusupov R.M. A model-oriented system for operational forecasting of river floods // Herald of the Russian Academy of Sciences, 89(4): 405–417, 2019. DOI: 10.1134/S1019331619040130.

8. Rostova, E., Rostov, N., Sobolevsky, V., Zakharov, V. Design and simulation of biotechnical multidimensional motion control systems of a robot manipulator // Proceedings - European Council for Modelling and Simulation, ECMS. 33(1). 2019. Стр. 145–150. DOI: 10.7148/2019-0145.

9. Sokolov, B., Mikoni, S., Sobolevsky, V., Zakharov, V., Rostova, E. Quality evaluation of models and polymodel complexes: Subject-object approach // Proceedings - European Council for Modelling and Simulation, ECMS. 2018. Стр. 305–310. DOI: 10.7148/2018-0305.

10. Petrovskiy, D., Barashkov, A., Sobolevsky, V., Sokolov, B., Pjatkov, V. On the real time logistics monitoring system development using artificial neural network // International Conference on Harbour, Maritime and Multimodal Logistics Modelling and Simulation. 2018. Стр. 14–20.

Патенты и свидетельства о регистрации программ для ЭВМ:

1. Соколов Б.В., Соколевский В.А. Программа автоматизированной генерации и обучения искусственных нейронных сетей. Свидетельство №2021668925. Зарегистрировано в реестре программ для ЭВМ 22.10.2021.

2. Соколевский В.А. Программа автоматизированного распознавания и подсчёта северных оленей на аэрофотоснимках. Свидетельство № 2022665074. Зарегистрировано в реестре программ для ЭВМ 09.08.2022.

Автореферат диссертации

СОБОЛЕВСКИЙ

Владислав Алексеевич

КОМПЛЕКСНАЯ АВТОМАТИЗАЦИЯ СИНТЕЗА ИСКУССТВЕННЫХ
НЕЙРОННЫХ СЕТЕЙ ПРЯМОГО РАСПРОСТРАНЕНИЯ

Текст автореферата размещен на сайтах:

Высшей аттестационной комиссии при Министерстве науки и высшего
образования Российской Федерации

<https://vak.minobrnauki.gov.ru/>

Федерального государственного бюджетного учреждения науки
«Санкт-Петербургский Федеральный исследовательский центр
Российской академии наук»

<http://www.spiiras.nw.ru/dissovet/>

Подписано в печать "29" ноября 2022 г.
Формат 60x84 1/16. Бумага офсетная. Печать офсетная.
Усл.печ.л. 1,0. Тираж 100 экз.
Заказ № ____