

БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ФАКУЛЬТЕТ РАДИОФИЗИКИ И КОМПЬЮТЕРНЫХ ТЕХНОЛОГИЙ
БИОЛОГИЧЕСКИЙ ФАКУЛЬТЕТ

КОМПЬЮТЕРНЫЕ ТЕХНОЛОГИИ И АНАЛИЗ ДАННЫХ (СТДА'2024)

**Материалы
IV Международной научно-практической
конференции**

Минск, 25–26 апреля 2024 г.

Научное электронное издание

МИНСК, БГУ, 2024

ISBN 978-985-881-636-0

© БГУ, 2024

УДК 004(06)+004.6(06)
ББК 32.97я431+32.973-018.2я

Редакционная коллегия:

кандидат физико-математических наук, доцент *В. В. Скакун* (гл. ред.);
кандидат физико-математических наук, доцент *Н. Н. Яцков*;
кандидат биологических наук, доцент *В. В. Гринёв*

Рецензенты:

кандидат технических наук, доцент *В. С. Садов*;
кандидат физико-математических наук, доцент *Е. И. Козлова*

Компьютерные технологии и анализ данных (СТДА'2024) [Электронный ресурс] : материалы IV Междунар. науч.-практ. конф., Минск, 25–26 апр. 2024 г. / Белорус. гос. ун-т ; редкол.: В. В. Скакун (гл. ред.), Н. Н. Яцков, В. В. Гринёв. – Минск : БГУ, 2024. – 1 электрон. опт. диск (CD-ROM). – ISBN 978-985-881-636-0.

Представлены новейшие результаты исследований в области компьютерных технологий, моделирования и анализа больших данных. Рассмотрены задачи анализа и понимания сложных и распределенных наборов физических и биологических данных. Особое внимание уделено вопросам вычислительного интеллекта, включающего компьютерное моделирование, интеллектуальные системы и программные технологии, искусственный интеллект, интеллектуальный анализ данных, машинное обучение и нейронные сети.

Минимальные системные требования:

PC, Pentium 4 или выше; RAM 1 Гб; Windows XP/7/10;
Adobe Acrobat

Оригинал-макет подготовлен в программе Microsoft Word

В авторской редакции

Ответственный за выпуск *В. В. Скакун*

Подписано к использованию 10.06.2024. Объем 2,65 МБ

Белорусский государственный университет.
Управление редакционно-издательской работы.
Пр. Независимости, 4, 220030, Минск.
Телефон: (017) 259-70-70.
e-mail: urir@bsu.by
<http://elib.bsu.by>

СОДЕРЖАНИЕ

ПРЕДИСЛОВИЕ	9

СЕКЦИЯ

«СИСТЕМЫ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ»

<i>Сюй С., Скакун В.В.</i> Исследование влияния нормализации данных на качество сегментации ядер клеток на иммуногистологических изображениях раковых клеток	11
<i>Жоголь Д.П., Лисица Е.В., Скакун В.В.</i> Анализ иммуногистохимических флуоресцентных изображений с помощью нейронных сетей глубокого обучения.....	15
<i>Семененко Л. В., Кочик Е. Н., Сорока А. М., Шкабара В. В.</i> Распознавание объектов на спутниковых снимках с использованием нейронных сетей: практика разработки и внедрения	19
<i>Лихимович Е.В., Лисица Е.В.</i> Разработка искусственной нейронной сети для анализа люминесцентных изображений раковых клеток.....	23
<i>Малый С.В., Орлова А.С.</i> Синтез структурно неоднородных блоков на основе технологий глубокого обучения	27
<i>Сюй С., Скакун В.В.</i> Применение метода главных компонент для формирования обучающего набора при сегментации ядер клеток раковых тканей с помощью нейронных сетей глубокого обучения	30
<i>Саечников И.В., Скакун В.В., Чернявская Э.А.</i> Распознавание и анализ морских судов методами глубокого обучения на базе аэрокосмических данных	34
<i>Халин А.А.</i> Проблема качества данных для машинного обучения.....	37

СЕКЦИЯ

«ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ»

<i>Qiu Shifeng, Sadov V.S.</i> Comparative study of STC and LSB image coding methods based on steganographic modification statistics.....	40
<i>Беляев Т. Е., Павлышко М. А., Баранова В. С., Николаев А. О.</i> Система ориентации инженерной модели наноспутника по солнцу.....	44
<i>Бородина М.Е., Кудинов В.А.</i> Технология обработки текста для дальнейшего семантического анализа.....	48
<i>Го Чжуню, Садов В.С.</i> Распознавание эмоционального состояния - «счастливый человек».....	52
<i>Гринько С.С., Индюков С.С., Лобанок М.В., Людчик О.Р.</i> Реализация учебного комплекса для измерения характеристик массивов лазерных пробоев в стекле.....	56

<i>Заблоцкая Д.М., Садов В.С.</i> Оптимизация алгоритма стеганографического скрывания служебной информации в пространственном представлении картографических изображений	59
<i>Камоцкий А.А., Гринько С.С., Семенович А.С., Лобанок М.В.</i> Система дистанционного мониторинга барометрических и метеорологических параметров окружающей среды	63
<i>Ким Т.Ю., Печковский Е.И., Печковская А.В.</i> Применение генетических алгоритмов для оптимизации компонентов редуктора при изготовлении с помощью трехмерной печати.....	67
<i>Кириленко А.О., Козлова Е.И.</i> Моделирование угроз устройствам интернета вещей.....	71
<i>Котова К.А., Садов В.С.</i> Алгоритм анализа микровыражений лица при выявлении лжи	75
<i>Лобанок М.В., Щербич И.Д., Семенович А.С., Павлышко М.А., Романов И.А.</i> Система дистанционного управления и анализа данных для исследования фотоприемников и фоточувствительных материалов	79
<i>Лобанок М.В., Зайков В.А., Гринкевич А.А., Щербич И.Д., Суконкина А.О.</i> Автоматизация измерений электрофизических характеристик тонкопленочных материалов	83
<i>Матусевич М.С., Гринько С.С., Скребец А.А., Лобанок М.В., Кадола Е.В.</i> Реализация стенда для измерений сигнальных характеристик лавинного фотодиода	87
<i>Paulenka D.A., Snezhko E.V., Kovalev V. A.</i> Testing and analyzing the performance of single-board computers for AI tasks.....	91
<i>Павлышко М.А., Семенович С.Н., Прохоров В.А.</i> Демодуляция и декодирование QPSK-радиосигнала стандарта CCSDS	96
<i>Павлышко М.А., Семенович С.Н., Семенович А.С., Суровцев М.А.</i> Стенд эмуляции ADS-B радиосигналов для тестирования модуля CMKA BSUSAT-2	99
<i>Прохоров В.А., Павлышко М.А., Никифоров А.В., Огурцов А.М., Шандицев А.А.</i> Реализация протокола передачи данных в бортовой сети сверхмалого космического аппарата.....	103
<i>Резниченко Д. В., Шалатонин И. А., Спиридонов А. А., Гранковская М. В.</i> Лабораторный стенд для мониторинга удаленных объектов с использованием протокола AX.25	106
<i>Рундыгин С. Д., Субач А. А., Труханович А. Л.</i> Разработка программного комплекса мониторинга и обработки данных по стандарту ISO 11898 для электрических транспортных средств.....	110
<i>Семенович А.С., Семенович С.Н., Стецко И.П.</i> Алгоритм для отладки процессов взаимной синхронизации узлов MESH-сети	114
<i>Семенович С.Н., Стецко И.П., Огурцов А.М., Терешко Д.Г., Калачёва А.В., Жилинская Т.А., Барсуков Е.А.</i> Функциональная структура инженерной модели сверхмалого космического аппарата.....	118
<i>Сидоренко А.В., Солодухо Н.А.</i> Алгоритм для огибания препятствий при движении мобильного робота	122

<i>Соболь А.М., Кочин В.П.</i> Инфраструктура умного кампуса	126
<i>Юшкевич Е.Д., Кураков И.А., Борейко Д. А., Ильюшенко А.С., Баранова В.С., Спиридонов А.А., Ушаков Д.В., Саечников В.А.</i> Разработка системы атмосферных зондов для организации клиент-серверной связи	129
<i>Ярмантович А.А., Гринадёр О.Д., Баранова В.С., Спиридонов А.А., Ушаков Д.В., Саечников В.А.</i> Разработка модуля оптической связи для межбортовой и межспутниковой коммуникации	133
<i>Семенович А.С., Семенович С.Н., Стецко И.П.</i> Реконфигурируемая аппаратно- программная платформа для сдачи нормативов комплекса ГТО	137
<i>Спиридонов А.А., Баранова В.С., Евчик В.Е., Шалатонин И. А., Черный В.Е., Ушаков Д.В., Саечников В.А.</i> Управление группировкой малоразмерных космических аппаратов распределенной сетью наземных станций	141
<i>Кочеткова П.А., Козлова Е.И.</i> Нейросетевое распознавание эмоций на статичных изображениях	145
<i>Шан Вэньли, Козлова Е.И.</i> Исследование взаимосвязи между вариабельностью сердечного ритма и выраженностью реакции вестибулярного аппарата человека при погружении в среду виртуальной реальности	149

СЕКЦИЯ

«КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ И СИСТЕМ»

<i>Борздов А.В., Борздов В.М., Петлицкий А.Н.</i> Моделирование фототока в кремниевых фотодиодах с PIN-структурой	153
<i>Вольская П.Д., Дещенко Г.Н.</i> Моделирование резонатора для измерения влажности зерна	157
<i>Гренкин Г.В., Карпо В.В., Чусова А.Е.</i> Ориентировка внутри зданий по упрощенному плану	160
<i>Жевняк О. Г., Борздов В. М., Борздов А. В.</i> Моделирование величины паразитного туннельного тока в элементах флеш-памяти	164
<i>Захаров М.Е., Ганджа Т.В.</i> Компьютерное моделирование теплообмника печи пиролиза для оптимизации процесса пиролиза углеводородов	168
<i>Стукач О.В., Каратиш Е.А.</i> Идентификация стохастической модели теплотребления в многоэтажных жилых зданиях города	169
<i>Ковалёв О.Ф., Шмигирёв Т.Ф.</i> Проектирование корпоративной информационной сети компании	176
<i>Кочин В.П.</i> Концептуальная модель и проектирование сложных интегрированных систем	180
<i>Мигель А.С., Малый С.В.</i> Исследование применимости метода связанных волн для расчета дифракционных спектров периодических решеток со сложным профилем	184
<i>Кугейко М.М., Лынько А.В.</i> Определение CO ₂ и паров воды в атмосфере	188

<i>Воротницкий Ю.И., Румас Р.А.</i> Модели однонаправленной передачи данных в компьютерных сетях	192
<i>Скакун В.В., Апанасович В.В.</i> Оценка концентрации молекул в нестационарной флуоресцентной флукуационной спектроскопии	196
<i>Тиманюк Д. В., Леонтьев А.В.</i> Онлайн-калькулятор для расчета выходных характеристик термодпар	200
<i>Ruiqi Huang, Bezrodny A.A.</i> Optimal control system structure formation of human-machine systems using genetical algorithms	203
<i>Шульган А.К., Воротницкий Ю.И.</i> Аппаратно-программная платформа исследования алгоритмов маршрутизации в беспроводной сенсорной сети.....	208
<i>Яцков Н.Н., Апанасович В.В., Яцков В.Н.</i> Генеративное имитационное моделирование сложных биофизических систем	211
<i>Яцков Н.Н., Апанасович В.В.</i> Нейросетевое имитационное моделирование при анализе экспериментальных данных флуоресцентной спектроскопии	215

СЕКЦИЯ «БИОИНФОРМАТИКА»

<i>Дигрис А.В., Вычик П.В., Дувалов Е.И., Скакун В.В., Николайчик Е.А.</i> Веб-инструмент для локального поиска сайтов связывания регуляторов транскрипции в последовательностях бактериальных геномов.....	219
<i>Chen Yu-Xiang, Tuzikov A.V.</i> A novel algorithm for prediction of drug resistance using cross-accuracy computation	223
<i>Воронова-Барте Н.В., Степаненко В.Б., Яхницкая Е.Р., Левыкина С.С.</i> Разработка программного окружения на основе Docker для обработки данных секвенирования	228
<i>Гузова Е.В., Ильющёнок И.Н., Барне Ф., Марин-Рубио Х.Л., Хайденрайх О., Гринев В.В.</i> Транскриптомно-протеомный подход к идентификации новых белков лейкозных клеток человека.....	232
<i>Dorofeev N.V., Sharapov R.V., Goryachev M.S.</i> Analysis of individual health changes by gait using a smartphone	236
<i>Kosareva A.A., Kovalev V.A., Fridman M.V.</i> Classification of whole-slide histopathological images of radiogenic papillary thyroid carcinoma.....	241
<i>Лайков Я.В., Тузиков А.В., Андрианов А.М.</i> Виртуальный скрининг потенциальных ингибиторов проникновения ВИЧ-1 с помощью методов молекулярного моделирования	246
<i>Левыкина С.С., Скамейко Д.С., Рабчун А.Ю., Шульгович А.М., Ракова М.Д., Воронова-Барте Н.В.</i> Характеристика UGT и GST генов в геноме тли <i>Macrosiphum albifrons</i>	250
<i>Левыкина С.С., Ракова М.Д., Воронова-Барте Н.В.</i> Сравнительный анализ ортогрупп внутри класса Насекомые	254

<i>Лиховец М.А., Гузова Е.В., Яцков Н.Н., Скакун В.В., Гринев В.В.</i> Идентификация структурных вариаций в геноме клеток острого миелоидного лейкоза по данным CAPTURE HI-C	258
<i>Мозоль Н.Р., Валентович Л.Н.</i> Разработка программного пакета для поиска смежных контигов при сборке геномов de novo	262
<i>Сарнацкий Д. Д., Яцков Н. Н., Гринев В. В.</i> Имитационная модель генерации сайтов однонуклеотидного полиморфизма в молекулах ДНК человека	265
<i>Саченко А.Б., Попичева Е.А., Щур В.В., Янцевич А.В.</i> Дизайн и анализ структуры гибридного белка ANGPT1_VEGF165 для лечения ишемии нижних конечностей	269
<i>Сиколенко М.А.</i> Прогнозирование регуляторных мишеней прокариотической РНК с помощью глубокого машинного обучения	272
<i>Трусов И.С., Гузова Е.В., Гринев В.В.</i> Биоинформатический пайплайн для оценки иммуногенности белков патогенов и раковых опухолей	276
<i>Чесалин В.И., Воронова-Барте Н.В., Левыкина С.С., Крук В.Л., Мешич Н.С., Александрович П.Е.</i> Идентификация совпадающих фрагментов некодирующих участков митохондриальных геномов насекомых отряда полужесткокрылые (лат. Hemiptera)	280
<i>Шаладонова М.И., Диченко Я.В., Усанов С.А.</i> Фармакофорная модель на основе структуры лиганда CYP17A1 человека	284
<i>Воронова-Барте Н.В., Левыкина С.С., Шульгович А.М.</i> Гены системы детоксикации в геноме тли <i>Macrosiphum rosae</i>	288

СЕКЦИЯ «ПРИКЛАДНОЙ АНАЛИЗ ДАННЫХ»

<i>Shuai Xiong, Яскевич В.Э.</i> Исследования электромагнитного канала утечки информации средств вычислительной техники	292
<i>Афоненко Ан. А., Афоненко А. А., Ушаков Д. В., Пушкарев С.С., Хабибуллин Р.А.</i> Анализ шероховатостей гетерограниц с использованием ПЭМ изображений квантово-каскадных гетероструктур	296
<i>Баранова В.С., Спиридонов А.А., Козлов В.Л., Ушаков Д.В., Саечников В.А.</i> Определение высоты орбиты космического объекта с использованием пассивных оптических измерений	300
<i>Глеба Е.М., Баранова В.С., Спиридонов А.А., Черный В.Е., Ушаков Д.В., Саечников В.А.</i> Методы оценки влияния космической погоды на бортовые системы низкоорбитальных спутников	304
<i>Кабулов Т.А., Кезик С.В., Недзьведь О.В., Мацкевич В.С., Демидчик В.В.</i> Автоматический анализ выявления зон корня на изображениях флуоресцентной микроскопии	308
<i>Колесников К. Д., Яцков Н. Н., Ковалев О. Ф.</i> Разработка локального алгоритма ограничения области поиска в задаче стереозрения	312

<i>Солодухо Н.А., Костенич А.А.</i> Разработка автоматического инвестиционного приложения с использованием он-лайн метрик	316
<i>Крамко Д.А., Садов В.С.</i> Интерфейс компьютерной программы «Прогнозирование психоневрологических исходов у недоношенных детей, родившихся с массой тела до 1500 грамм».....	320
<i>Крупович Д.В., Павлышко М.А., Камоцкий А.А.</i> Бортовой модуль для тестирования инженерной модели СМКА	324
<i>Лебедевский А.В., Фираго В.А., Левкович Н.В., Шулико К.И., Слодинская В.М.</i> Оперативное тестирование микросхем АЦП	328
<i>Фираго В.А., Левкович Н.В., Лебедевский А.В., Славинский С. Г., Шулико К.И.</i> Двухволновая рефракционная пульсоксиметрия.....	332
<i>Lisitsa Y.U.</i> Implementation of the watershed method for separating nuclei of cancer cells in fluorescent images	336
<i>Ромашко А.В., Яцков Н. Н.</i> Исследование методов снижения размерности в ходе анализа данных флуоресцентной спектроскопии	341
<i>Саечников А.В., Чернявская Э.А., Саечников В.А.</i> Многоканальный интеллектуальный иммуноферментный анализ на основе оптических микрорезонаторов.....	345
<i>Лебедевский А.В., Фираго В.А., Кугейко М.М., Левкович Н.В., Славинский С. Г., Шулико К.И.</i> Определение оксигинации крови коры головного мозга	348
<i>Фираго В.А., Лебедевский А.В., Левкович Н.В., Славинский С. Г., Шулико К.И., Гвоздев Н.С., Панченко Г.В.</i> Неинвазивная оценка состояния артериальных сосудов микроциркуляторного русла.....	352
<i>Хоменко С. Ю., Дигрис А.В.</i> Веб-приложение для анализа данных флуоресцентной спектроскопии	356

ПРЕДИСЛОВИЕ

Стремительное развитие компьютерных технологий, вычислительных систем, экспериментального оборудования и облачных хранилищ информации дало толчок активному внедрению методов анализа данных для создания высокоэффективных систем обеспечения экспериментальных исследований. В числе основных задач анализа данных следует отметить разработку перспективных методов анализа и моделирования, интерпретации и хранения информации, создание инфраструктуры для приема, обработки и распространения результатов экспериментальных исследований.

В сборнике материалов конференции «Компьютерные технологии и анализ данных» (CTDA'2024) представлены новейшие результаты в области компьютерных технологий, моделирования и анализа больших данных. Цель конференции – распространение теоретических и практических достижений в разработке моделей, методов и программных средств для обработки больших данных, полученных в физике, информатике, биоинформатике, инженерии и прочих направлениях практической деятельности. Рассмотрены задачи анализа и понимания сложных и распределенных наборов физических и биологических данных. Особое внимание уделено вопросам вычислительного интеллекта, включающего компьютерное моделирование, интеллектуальные системы и программные технологии, искусственный интеллект, интеллектуальный анализ данных, машинное обучение и нейронные сети.

Международная научно-практическая конференция «Компьютерные технологии и анализ данных» (CTDA'2024) состоялась 25–26 апреля 2024 года в Белорусском государственном университете на факультете радиоп физики и компьютерных технологий при поддержке биологического факультета. Научный программный комитет конференции сформирован из ведущих специалистов университетов и научных центров Белоруссии и России. В работе конференции приняли участие 90 авторов из многих университетов и научно-производственных компаний из Республики Беларусь и Российской Федерации.

Тематика конференции включала следующие направления:

- *системы машинного и глубокого обучения* – передовые аналитические системы обработки информации в области искусственного интеллекта, обеспечивающие автоматизацию и высокую эффективность анализа и обработки данных;
- *интеллектуальные технологии и системы* – направление включает разработку и применение методов искусственного интеллекта для создания систем, способных на основе использования интеллектуальных технологий проектирования, алгоритмизации, анализа и хранения данных

(информации) целенаправленно приобретать, воспроизводить и использовать знания или выполнять задачи, требующие человеческого интеллекта;

- *компьютерное моделирование процессов и систем* – использует математические и компьютерные методы для имитации реальных объектов и явлений; позволяет проводить эксперименты и исследования в виртуальной среде, что является одним из самых эффективных инструментов для анализа, изучения и проектирования различных процессов и систем;

- *прикладной анализ данных* – направление, включающее методы, алгоритмы и программные средства для обработки, анализа и интерпретации данных при решении конкретных задач прикладного анализа данных;

- *биоинформатика* – направление междисциплинарных исследований, включающее разработку математических моделей, алгоритмов и программных средств для анализа и интерпретации больших биологических и биомедицинских данных.

В Белорусском государственном университете на протяжении многих лет проводятся семинары по указанным направлениям с участием сотрудников, студентов и аспирантов, которые и выступают с инициативой организации данной конференции. Оргкомитет Международной научно-практической конференции «Компьютерные технологии и анализ данных» (СТДА'2024) убежден в том, что сборник материалов конференции способствует постоянному обмену опытом и эффективному сотрудничеству специалистов из университетов Республики Беларусь, Российской Федерации, Китайской Народной Республики и других стран.

СЕКЦИЯ «СИСТЕМЫ МАШИННОГО И ГЛУБОКОГО ОБУЧЕНИЯ»

ИССЛЕДОВАНИЕ ВЛИЯНИЯ НОРМАЛИЗАЦИИ ДАННЫХ НА КАЧЕСТВО СЕГМЕНТАЦИИ ЯДЕР КЛЕТОК НА ИММУНОГИСТОЛОГИЧЕСКИХ ИЗОБРАЖЕНИЯХ РАКОВЫХ КЛЕТОК

С. Сюй, В.В. Скакун

Белорусский государственный университет, Минск, Беларусь;

E-mail: xusilun@hotmail.com

Сегментация ядер клеток на гистологических изображениях срезов раковых тканей позволяет оценить степень развития заболевания. Основным инструментом для выполнения сегментации являются нейронные сети глубокого обучения. В данной работе произведено исследование применения нормализации данных на качество сегментации. Наилучший результат был получен при использовании нормализации FRN и случайного выбора патчей при формировании пакетов.

Ключевые слова: глубокое обучение; изображения раковых клеток; сегментация изображений; нормализация.

ВВЕДЕНИЕ

В медицинской диагностике раковых заболеваний важное место занимает анализ иммуногистологических изображений раковых клеток. Сегментация ядер раковых клеток позволяет произвести подсчет количества пораженных клеток и таким образом оценить степень заболевания. В настоящее время основным инструментом сегментации изображений, в том числе иммуногистологических, являются сверточные нейронные сети (англ. Convolutional Neural Network, CNN).

Процесс обучения практически любой нейронной сети критически зависит от качества входных данных. Особенно это заметно, когда объем обучающей выборки невелик. Увеличить объем выборки можно путем разделения исходного иммуногистологического изображения на несколько частей (патчей) с перекрытием и применения аугментации [1]. Вследствие большой гетерогенности изображений раковых клеток разброс значений пикселей в каждом из патчей может существенно меняться. Для нормализации входных данных применяются разные подходы. Традиционными являются преобразование значений пикселей к диапазону $[0; 1]$ или просто путем деления значений пикселей на их максимально возможное значение (255 для 8-битного изображения) и стандартизация изображений путем деления значений пикселей на их среднеквадратическое

отклонение с последующим вычитанием среднего. Дополнительно, патчи небольшого размера могут существенно отличаться друг от друга, будучи выбранными из области со здоровыми или больными клетками.

Процесс обучения нейронной сети по своей сути заключается в изучении распределения данных. Сеть состоит из множества слоев. После каждого слоя разброс значений пикселей неизбежно меняется. Если данные не нормализованы, распределение каждой партии обучающих данных будет другим. Нейронной сети приходится находить точку баланса между этими многочисленными распределениями. Таким образом, нейронной сети сложно сходиться. Если мы просто нормализуем входные данные, мы сможем гарантировать только то, что распределение данных на входе будет одинаковым, но не на всех уровнях сети. Поэтому нормализацию данных желательно производить после выполнения каждой свертки. Нормализация может также действовать как форма регуляризации данных, снижая риск переобучения и повышая способность модели к обобщению.

В настоящее время существует несколько методов нормализации, применяемых в CNN: пакетная нормализация (Batch Normalization, BN) [2], нормализация отклика фильтра (Filter Response Normalization, FRN) [3]. Существуют и другие виды нормализации, применяемые в рекуррентных и генеративных сетях. Пакетная нормализация плохо работает при малом размере пакета (Batch Size). Нормализация отклика фильтра специально разрабатывалась для устранения влияния размера пакета на нормализацию [3]. Выбор вида нормализации в каждом конкретном случае не очевиден. Соответственно, целью работы является исследование влияния нормализации данных на качество сегментации ядер клеток на иммуногистологических изображениях раковых клеток.

РЕЗУЛЬТАТЫ

Исходными данными в данной работе являются 9 флуоресцентных изображений срезов ткани опухоли молочной железы в формате RGB, полученные при 10-кратном увеличении с использованием эпифлуоресцентного инвертированного микроскопа Nikon TE200, оснащенного ПЗС-камерой Photometrics серии 300 [4]. Изображения 3-х каналные, 2048x2048 пикселей. Синий краситель 4,6-диамидино-2-фенилиндола дигидрохлорид (DAPI) использовался для маркировки всех ядер. Красный цианиновый краситель Cy5 использовался для маркировки эстрогена рецептора, содержащегося в основном в ядрах раковых клеток. Белок цитокератин, накапливающийся в цитоплазме раковых клеток, был помечен зеленым цианиновым красителем Cy3. Размеченные экспертами изображения, на кото-

рых выделены только маски ядер раковых клеток (ground truth, GT), представляют собой бинарные изображения, в которых пикселям ядер клеток присвоено значение 1 [4].

В процессе предварительной подготовки данных, заключающейся в выделении областей размером 256×256 пикселей с перекрытием наполовину как по вертикале, так и по горизонтали с последующим выделением из каждой полученной области по 8 патчей размером 128×128 пикселей со случайным смещением, получено 12064 патча размером 128×128 [1]. Изображение одного патча представлено на рис. 1.

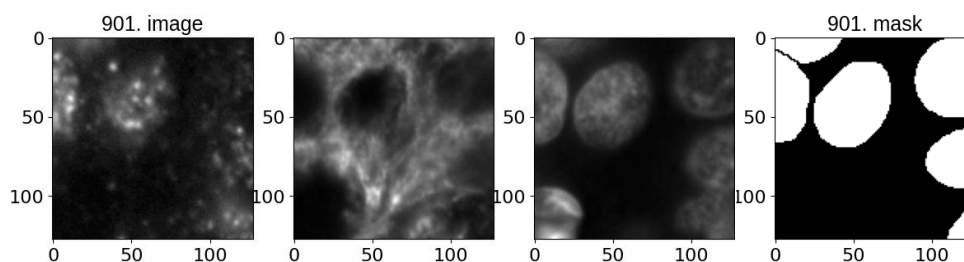


Рис. 1. Фрагмент иммуногистологического флуоресцентного изображения раковых клеток размером 128×128 пикселей. Слева направо: красный канал, зеленый канал, синий канал, размеченные маски ядер.

Для исследования влияния нормализации на качество сегментации ядер использовалась модель U-Net. Архитектура сети и ее параметры подробно описаны в работе [1]. Все тесты проводились в среде Google Colab, где аппаратная среда – графический процессор V100 (16 ГБ) и оперативная память 51 ГБ. Нейронная сеть реализована на основе фреймворка TensorFlow2. Параметры исследования следующие: BatchSize равен 32, количество итераций 30, оптимизатор Adam. Для оценки качества сегментации выполнялся расчет двух метрик: бинарное пересечение над объединением на уровне 0,5 (BIOU) и функция бинарной энтропии в качестве функции потерь. Исходные данные были разбиты на обучающую, валидационную и тестовую выборки в следующих количествах: 9384, 1340, 1340.

Исследовалось влияние нормализации как входных данных, так и данных в процессе обучения. Вследствие большой гетерогенности изображений раковых клеток на вход могут подаваться вначале серии патчей больных, потом здоровых, по своему виду сильно отличающихся клеток. Для исключения влияния вида клеток логичным является формирование пакетов данных для обучения путем случайной выборки патчей. Соответственно, исследовались следующие варианты: 1) стандартная сеть U-Net без применения нормализации; 2) добавление слоев BN после каждого слоя свертки в энкодере и декодере; 3) добавление слоев FRN после каждого слоя свертки в энкодере и декодере; 4) вариант 2) со случайным выбором патчей; 5) вариант 3) со случайным выбором патчей. Наилучший

результат был получен для варианта 5), наихудший – для варианта 1). Как видно из диаграммы, разброс значений метрики BIoU для тестовой и валидационной выборок существенно меньше для случаев 4) и 5).

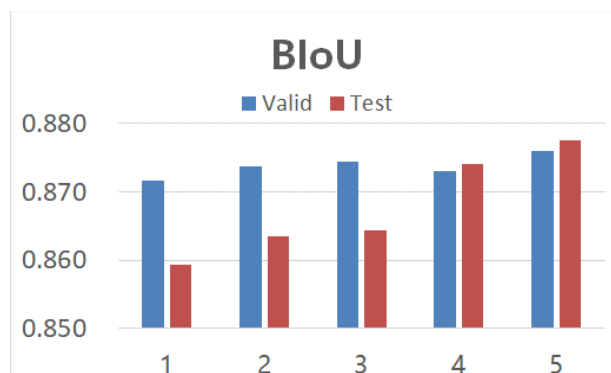


Рис. 2. Значения метрики BIoU при исследовании влияния нормализации данных на качество сегментации. Синим цветом отображен показатель BIoU, полученный на валидационном наборе; красным – на тестовом наборе.

ВЫВОДЫ

В этом исследовании мы сравнили влияние нормализации данных на качество сегментации ядер на иммуногистологических флуоресцентных изображениях раковых клеток молочной железы. Добавления слоев нормализации улучшает качество сегментации и стабилизирует процесс обучения. Применение нормализации функции отклика FRN показало наилучший результат при сегментации ядер раковых клеток на флуоресцентных изображениях. Формирование обучающей выборки (пакетов) путем случайного выбора патчей из всего набора данных позволило дополнительно улучшить качество сегментации и выровнять показатели метрик, вычисляемые на основе валидационной и тестовой выборок, что свидетельствует об улучшении способности модели к обобщению.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Xu Silun, Skakun V. Comparison of Deep Learning Preprocessing Algorithms of Nuclei Segmentation on Fluorescence Immunohistology Images of Cancer Cells // *PRIP*. 2022. P. 166-177. DOI:10.1007/978-3-030-98883-8_12.
2. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // arXiv:1502.03167. 2015.
3. Singh S., Krishnan S. Filter Response Normalization Layer: Eliminating Batch Dependence in the Training of Deep Neural Networks // arXiv:1911.09737v2. 2019.
4. Lisitsa Y.V., Yatskou M.M., Apanasovich V.V., Apanasovich T.V. Algorithm for automatic segmentation of nuclear boundaries in cancer cells in three-channel luminescent images // *Journal of Applied Spectroscopy* 82(4). 2015. P. 598-607

АНАЛИЗ ИММУНОГИСТОХИМИЧЕСКИХ ФЛУОРЕСЦЕНТНЫХ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ

Д.П. Жоголь, Е.В. Лисица, В.В. Скакун

Белорусский государственный университет, Минск, Беларусь

E-mail: danzhog@gmail.com

В данной работе описываются результаты сегментации иммуногистохимических флуоресцентных изображений раковых клеток молочной железы с помощью нейронной сети глубокого обучения Detectron 2. Нейронная сеть показала хорошие результаты сегментирования ядер раковых клеток с автоматическим выделением их в отдельные объекты, одновременно решая сложную задачу разделения слипшихся ядер.

Ключевые слова: Detectron2; глубокое обучение; сегментация изображений; сегментация объектов; иммуногистохимические изображения.

ВВЕДЕНИЕ

Гистология получила широкое применение в диагностической медицине при исследовании онкологических заболеваний. На 250 тысяч случаев МРТ диагностики приходится около полутора миллионов иммуногистохимических исследований. Однако проведение иммуногистохимических исследований по-прежнему затруднено ручным анализом данных, который требует как подготовки специалистов, так и временных ресурсов для визуального анализа. Необходимость в автоматизированных системах для данного процесса остается актуальной. Одним из методов количественной оценки онкологических процессов является подсчет количества пораженных клеток на изображениях. Соответственно, требуется выделить области, пораженные заболеванием, и в них произвести сегментацию раковых клеток. Однако, сложности с рутинной настройкой параметров и обработкой изображений препятствуют проведению анализа в приемлемые сроки.

На сегодняшний момент имеется значительное количество архитектур нейронных сетей, успешно применяемых для сегментации ядер раковых клеток [1]. Однако, одной из основных проблем, связанных с использованием сверточных нейронных сетей для сегментации изображений, является их ограничение в осуществлении семантической сегментации, т.е. определения метки принадлежности каждого пикселя к определенной категории. Это влечет за собой невозможность прямого выделения объектов изображения. Для решения данной проблемы требуется сложная постобработка результатов сегментации, включающая разделение "слипшихся" объектов и удаление неправильно сегментированных.

Сравнительно недавно был анонсирован фреймворк нового поколения Detectron 2, разработанный исследовательской группой Facebook AI Research [2]. Detectron 2 основан на фреймворке PyTorch и обеспечивает единый интерфейс при решении разнообразных задач анализа изображений, включая обнаружение объектов, сегментацию экземпляров и паноптическую сегментацию. Он включает в себя набор заранее обученных моделей машинного обучения, таких как Faster R-CNN [3], Mask R-CNN [4], RetinaNet [5] и DensePose [6], для обнаружения объектов и сегментации сцен. Цель данного исследования заключается в оценке эффективности применения фреймворка Detectron2 для сегментации ядер клеток на иммуногистохимических изображениях срезов раковых тканей.

РЕЗУЛЬТАТЫ

В статье исследуются трехканальные флуоресцентные изображения раковых клеток молочной железы, полученные с 10-кратным увеличением на инвертированном эпифлуоресцентном микроскопе Nikon TE200, оснащенном CCD-камерой серии Photometrics 300. Размер изображений 2048×2048 пикселей. Индикатором раковых клеток является белок эстроген-рецептор, помеченный красителем Cy5 (регистрируемый в красном канале изображения). В цитоплазме раковых клеток накапливается белок цитокератин, маркированный красителем Cy3 (регистрируемый в зеленом канале). Для маркировки всех ядер использован краситель 4,6- DAPI (регистрируемый в синем канале). Соответственно, маркерами раковых клеток являются два красителя – Cy5 и Cy3. В распоряжении авторов имеется 9 экспериментальных изображений, вручную размеченных экспертами. В размеченных бинарных изображениях маски ядер раковых клеток имеют значение 1.

Каждое исходное изображение было разделено на 16 фрагментов размером 512×512 пикселей, что дало в общей сложности 144 изображения. Эти изображения были далее разделены на наборы данных для обучения, валидации и тестирования, состоящие из 109, 20 и 15 изображений соответственно. Каждому изображению приписана бинарная маска, которая была преобразована в формат COCO. Формат COCO представляет собой JSON-файл, содержащий информацию об изображении, включая путь к файлу, его размеры, а также список аннотированных объектов, каждый из которых представлен ограничивающим прямоугольником и классом объекта. Далее изображения были нормализованы с применением метода, предложенного в исследовании [7].

Для исследования применения фреймворка Detectron2 для сегментации ядер раковых клеток был выбран язык программирования Python и библиотеки Detectron2, PyTorch, OpenCV, scikit-image, seaborn, pandas,

NumPy, matplotlib. Разработка велась в среде Spyder. Реализованная архитектура нейронной сети условно разделяется на CNN-сеть вычисления признаков изображения, называемую backbone, и сеть head, являющуюся результатом работы отдельных нейронных сетей, отвечающих за предсказание охватывающей рамки, классификацию объекта и определение его маски. Функция потерь для них общая и включает три компонента:

$$L_{all} = L_{cls} + L_{box} + L_{mask}, \quad (1)$$

где L_{all} – общая функция потерь модели; L_{cls} – функция потерь классификации (перекрестная энтропия между предсказанными и истинными классами объектов); L_{box} – функция потерь ограничивающей рамки; L_{mask} (mask loss) – функция потерь маски (бинарная перекрестная энтропия для каждого пикселя). Для оценки эффективности рассчитывались метрика точности (Accuracy):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}, \quad (2)$$

где TP (true positive) – верно идентифицированные объекты класса, TN (true negative) – верно отклоненные объекты, FP (false positive) – неверно идентифицированные объекты (как принадлежащие классу), FN (false negative) – неверно отклоненные объекты (как принадлежащие классу), а также метрика средней точности (Average Precision, AP):

$$AP = \int_0^1 p(r) dr \quad (3)$$

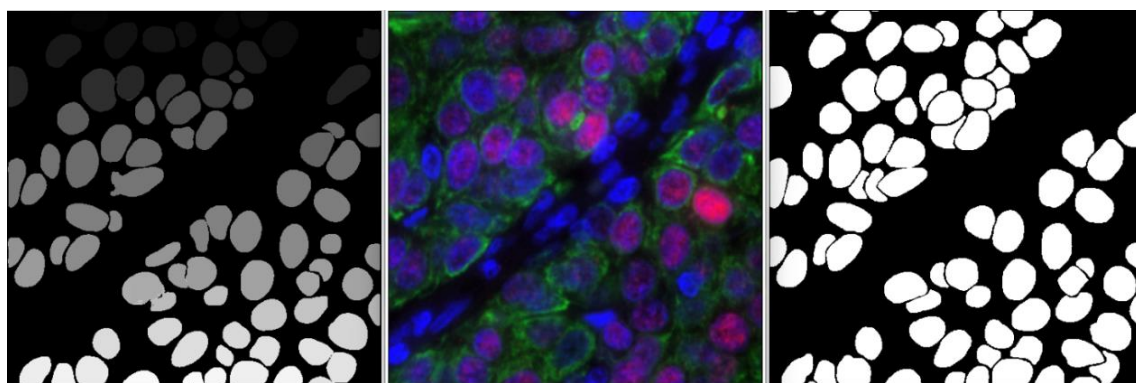
где p – точность ($Precision = \frac{TP}{TP + FP}$), r – полнота ($Recall = \frac{TP}{TP + FN}$) из предположения, что ограничивающая рамка определена верно, если пересечение над объединением (Intersection over Union) $IoU \geq 0,5$.

Размер пакета (Batch size) установлен в 2. Скорость обучения оптимизатора $2,5e-4$. Обучалась модель в течении 10000 итераций. Вычисления производились с помощью видеоускорителя Nvidia GeForce RTX 1070 с 8 ГБ видеопамати. Обучение в сумме заняло около 2 часов, что показывает высокую обучаемость нейронной сети.

Полученные значения метрик приведены в таблице. AP50 обозначает Average Precision при пороге $IoU = 0.5$. Это означает, что предсказанная ограничивающая рамка считается верной, если её пересечение с истинной ограничивающей рамкой (IoU) больше или равно 50%. AP75 обозначает Average Precision при более строгом пороге IoU равном 0.75. Результаты сегментации одного фрагмента представлены на рисунке.

Показатели метрик на тестовом наборе данных

AP	AP50	AP75	Accuracy
63,236	87,278	74,315	0,923



Результаты сегментации. Слева – результат сегментации изображения с помощью Detectron 2, представленный 16-битной маской. В центре – исходное изображение. Справа бинарная маска (ground truth).

ВЫВОДЫ

Применение фреймворка Detectron2 для задачи сегментации ядер клеток на иммуногистохимических изображениях продемонстрировало обнадеживающие результаты. Нейронная сеть была обучена на малой выборке из 109 размеченных изображений размером 512×512 . Главное преимущество данного фреймворка заключается в возможности выделения каждого ядра как отдельного объекта, что значительно упрощает пост-обработку полученных результатов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. A review of deep-learning-based medical image segmentation methods / Liu X. [et al.] // Sustainability. V. 13. №. 3. 2021. P. 1224.
2. Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo and Ross Girshick. Detectron 2. [Электронный ресурс]. URL: <https://github.com/facebookresearch/detectron2> (дата обращения: 01.03.2024).
3. Faster R-cnn: Towards real-time object detection with region proposal networks / Ren S. [et al.] // Advances in neural information processing systems. V. 28. 2015.
4. Mask R-cnn / He K. [et al.] // Proceedings of the IEEE international conference on computer vision. 2017. P. 2961-2969.
5. Focal loss for dense object detection / Lin T. Y. [et al.] // Proceedings of the IEEE international conference on computer vision. 2017. P. 2980-2988.
6. Güler R. A., Neverova N., Kokkinos I. Densepose. Dense human pose estimation in the wild // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. P. 7297-7306.
7. A method for normalizing histology slides for quantitative analysis / Macenko M. [et al.] // IEEE international symposium on biomedical imaging: from nano to macro. IEEE. 2009. P. 1107-1110.

РАСПОЗНАВАНИЕ ОБЪЕКТОВ НА СПУТНИКОВЫХ СНИМКАХ С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ: ПРАКТИКА РАЗРАБОТКИ И ВНЕДРЕНИЯ

Л. В. Семененко, Е. Н. Кочик, А. М. Сорока, В. В. Шкабара

*Научно-исследовательское учреждение «Институт прикладных физических проблем
имени А. Н. Севченко» Белорусского государственного университета (НИИПФП им.
А.Н. Севченко БГУ), Минск, Беларусь
eugene.kochik@gmail.com*

Рассмотрена комплексная методика организации процесса обучения нейронной сети для автоматизации экологического мониторинга. Приведена практика использования данной методики на примере создания информационно-аналитической системы контроля и анализа деятельности в водоохранных зонах.

Ключевые слова: экологический мониторинг; нейронные сети; ДЗЗ; БКСДЗ.

ВВЕДЕНИЕ

Уже много лет для контроля и анализа деятельности в природопользовании используются спутниковые данные. Системы искусственного интеллекта, в первую очередь нейронные сети глубокого обучения, в последние годы открыли широкий спектр новых возможностей по использованию спутниковых данных.

Следует отметить, что большинство разработок в этой области на сегодняшний день носят научно-исследовательский характер и их практическое использование часто представляется проблематичным.

Коллектив лаборатории информационных технологий НИИПФП им. А.Н. Севченко БГУ автоматизировал для Министерства природных ресурсов и охраны окружающей среды Республики Беларусь процессы экологического мониторинга на основе спутниковых снимков с использованием нейронных сетей глубокого обучения в рамках проектов по созданию информационно-аналитической системы поддержки принятия решений по контролю и анализу деятельности по добыче полезных ископаемых и информационно-аналитической системы контроля и анализа деятельности в водоохранных зонах (ИАС «Водо-охранные зоны»).

В процессе разработки нами был организован поэтапный комплекс работ по обучению нейронных сетей, который предлагается рассмотреть ниже на примере разработки ИАС «Водоохранные зоны».

Комплекс работ включал в себя три больших этапа:

1. Аналитическое изучение предметной области и постановка задачи;
2. Итерационное обучение нейронной сети;
3. Интеграция обученной модели в процесс мониторинга.

ПОСТАНОВКА ЗАДАЧИ НА КОНТРОЛЬ ХОЗЯЙСТВЕННОЙ ДЕЯТЕЛЬНОСТИ В ВОДООХРАННЫХ ЗОНАХ С ИСПОЛЬЗОВАНИЕ СПУТНИКОВЫХ ДАННЫХ И НЕЙРОННЫХ СЕТЕЙ

В связи с трудоёмкостью наземного мониторинга актуальной задачей является контроль хозяйственной деятельности в водоохранных зонах с использованием спутниковых снимков, предоставляемых на безвозмездной основе, и нейронных сетей.

Спутниковые данные и нейронные сети имеют ряд ограничений и не могут обеспечить качественного детектирования всех объектов. Поэтому на первом этапе необходимо разработать четкую постановку задачи.

При реализации проекта постановка задачи проводилась пошагово.

На первом шаге проведено изучение систем дистанционного зондирования Земли (далее – ДЗЗ), которые могли быть использованы для контроля хозяйственной деятельности в природных системах. Перспективными были признаны Белорусская космическая система дистанционного зондирования Земли (далее – БКСДЗ) и Sentinel-2.

На втором шаге проведен анализ нормативных документов на предмет выделения состава возможных нарушений и возможностей их контроля на основе спутниковых данных. Был определен предварительный перечень контролируемых объектов.

На третьем шаге проведен анализ актуальности и перспективности нарушений из предварительного перечня. Было проведено визуальное дешифрирование территории водоохранных зон 8-ми районов Минской области на снимках сверхвысокого разрешения. Затем – распознавание обнаруженных объектов на снимках БКСДЗ и Sentinel-2. Принадлежность обнаруженного объекта к виду хозяйственной деятельности определялась на основе информационных ресурсов Государственного комитета по имуществу Республики Беларусь. Анализ позволил сформулировать итоговую задачу, которой стало выявление новых объектов загрязнения на спутниковых данных БКСДЗ. Это важная задача, которая позволит оптимизировать процесс разработки необходимых природоохранных мероприятий [1].

На четвертом шаге проводилось изучение процесса разработки и актуализации проектов водоохранных зон и прибрежных полос.

На пятом шаге проведено изучение объектов загрязнения. По результатам определен актуальный объект контроля: вновь возникшие здания и сооружения площадью свыше 6 ар, а также места проведения строительных работ.

ИТЕРАЦИОННОЕ ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ

Задача появления новых объектов загрязнения в водоохранных зонах решается на основе контроля изменений на парах разновременных спутниковых снимков. Предобученные модели нейронных сетей не обеспечивают качественного детектирования на спутниковых снимках БКСДЗ. Поэтому для решения задачи проводилось обучение отобранных моделей нейронных сетей с формированием обучающей выборки. Формальная постановка задачи – семантическая сегментация изображений на два класса: «нет изменений», «есть изменения». В качестве итогового решения была выбрана нейронная сеть SegNeXt [2].

На первом этапе формирования обучающей выборки была реализована разметка на паре снимков за 2019 и 2021 годы. В разметку вошли все объекты интереса и ложные объекты, похожие на объекты интереса по яркостным и семантическим характеристикам (7 типов).

Далее обучающая выборка накапливалась итерационно. Размер выборки на каждом этапе приведен в таблице ниже.

Размер обучающей выборки

	Этап 1	Этап 2	Этап 3	Этап 4	Этап 5	Этап 6	Этап 7
Количество	2445	4005	5365	6550	10027	12114	12798

Между итерациями проводилось обучение нейронной сети с изменением параметров, находились дефекты в протоколе разметки, обновлялась валидационная выборка.

Для оценки итогового качества была использована метрика mIoU, которая рассчитывалась как среднее значение метрики IoU (Intersection over Union) для обоих классов, где сама IoU рассчитывается как отношение площади пересечения двух областей (эталонной и текущей) к общей площади этих областей [3]. Формально, для двух непустых множеств A и B, функция IoU определяется как:

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|};$$

Итоговое значение метрики mIoU – 0,7576.

РЕАЛИЗАЦИЯ ПРОЦЕССА МОНИТОРИНГА С ИСПОЛЬЗОВАНИЕМ СПУТНИКОВЫХ ДАННЫХ И НЕЙРОННОЙ СЕТИ

Обученная модель нейронной сети была встроена в комплекс программных средств обработки данных ДЗЗ ИАС «Водоохранные зоны», который обеспечил автоматизацию следующих процессов:

1. Получение спутниковых данных с FTP-сервера УП «Геоинформационные системы»;
2. Обработка данных нейросетью и формирование заданий на проверку;
3. Заверка заданий оператором ДЗЗ;
4. Анализ сигналов потенциального объекта загрязнения.

Данный комплекс предоставил сотрудникам природоохранного ведомства возможности быстрого анализа появления новых потенциальных объектов загрязнения.

По итогам разработки ИАС «Водоохранные зоны» была проведена обработка снимков БКСДЗ за период с 22.07.2022 по 1.06.2023 (всего 27 снимков, в сумме покрывающих более 80% территории Республики Беларусь).

Нейронной сетью было детектировано 2949 потенциальных объектов нарушения, из них истинными оказалось 119, обладающих схожими с объектами интереса дешифровочными признаками - 1474.

ЗАКЛЮЧЕНИЕ

Наш опыт использования нейронных сетей глубокого обучения по детектированию объектов на данных ДЗЗ на реальном проекте показал, что обучение модели следует организовывать в виде комплексного поэтапного процесса, в котором особое внимание следует уделить формированию четкой постановки задачи. Обучение следует начинать на небольшой обучающей выборке и проводить ее итерационное наращивание, совместно с которым уточнять и корректировать протокол разметки, проводить оптимизацию параметров нейронной сети.

Часто используемые метрики качества на основании валидационной выборки не гарантируют соответствующего качества в промышленной эксплуатации, в связи с чем необходима разработка метрики, обеспечивающей комплексную оценку работы нейронной сети для автоматизируемого процесса мониторинга.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Гертман Л. Н., Буко И. Ю., Шариков А. П. Из опыта разработки проектов водоохраных зон и прибрежных полос в соответствии с действующим законодательством // Земля Беларуси. 2018. № 2. С. 20-23.
2. SegNeXt: Rethinking Convolutional Attention Design for Semantic Segmentation [Электронный ресурс]. URL: <https://arxiv.org/abs/1511.00561> (дата обращения: 12.09.2023). DOI: 10.48550/arXiv.1511.00561
3. Intersection over Union (IoU) [Электронный ресурс]. URL: <https://wiki.cloudfactory.com/docs/mp-wiki/metrics/iou-intersection-over-union> (дата обращения: 22.03.2023).

РАЗРАБОТКА ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ АНАЛИЗА ЛЮМИНЕСЦЕНТНЫХ ИЗОБРАЖЕНИЙ РАКОВЫХ КЛЕТОК

Е.В. Лихимович, Е.В. Лисица

Белорусский государственный университет, Минск E-mail:
elihimovic@gmail.com

Сегментация является важным инструментом в области компьютерного зрения и обработки изображений. В работе рассмотрена сегментация изображений раковых клеток с помощью нейронной сети UNet.

Ключевые слова: нейронная сеть; сегментация изображений; UNet.

ВВЕДЕНИЕ

Искусственные нейронные сети (ИНС) получили широкое применение в области сегментации изображения. Процесс сегментации представляет собой разделение изображений на более мелкие компоненты или объекты с целью более глубокого понимания их содержания и структуры. Наиболее мощным инструментом для сегментации являются сверточные нейронные сети, способные автоматически извлекать признаки из изображений. Одной из наиболее распространенных задач сегментации является семантическая сегментация, при которой каждый пиксель изображения классифицируется в определенный класс или категорию [1].

Нейронные сети широко используются в медицине, в частности в онкологии при диагностировании онкологических заболеваний. ИНС расширяют клинические возможности в диагностике и терапии злокачественных новообразований. Применение UNet в медицине имеет ряд преимуществ. Данная архитектура способствует автоматическому выделению и сегментации интересующих областей на изображениях, что представляет значимость для процессов диагностики и разработки лечебных стратегий. Некоторые конкретные области, в которых UNet применяется в медицине, включают сегментацию органов, обнаружение аномалий, сегментацию клеток и тканей, регистрацию изображений.

UNET

UNet представляет из себя симметричный сверточный автоэнкодер [1], где сверточные слои энкодера используются для извлечения признаков и сжатия информации, а транспонированные сверточные слои декодера выполняют восстановление и увеличение пространственного разрешения.

Основные компоненты UNet включают:

Энкодер – состоит из серии сверточных слоев, которые последовательно уменьшают пространственное разрешение изображения и повышают абстрактность признаков.

Декодер – этот блок выполняет обратную операцию энкодера, увеличивая пространственное разрешение и восстанавливая детали сегментированного изображения.

Подключения пропуска – связывают слои энкодера с соответствующими слоями декодера, что позволяет передавать более низкоуровневые детали и признаки из энкодера в декодер, что помогает сохранить информацию о местоположении и предотвращает потерю деталей в процессе декодирования.

ПОДГОТОВКА ДАННЫХ

Исходный набор данных состоял из 9 изображений и соответствующих им 9 масок раковых клеток. Для увеличения объема обучающей выборки изображения были разбиты на 16 частей путем горизонтального и вертикального разреза. Кроме того, данные были аугментированы с применением функции из пакета keras - `imagedatagenerator`, которая случайным образом применяет операции поворота, отражения по горизонтали и вертикали, а также сдвиги по горизонтали и вертикали к изображениям. После аугментации и разбиения набор данных состоял из 5000 изображений и их масок (рис. 1).

ОБУЧЕНИЕ МОДЕЛИ

ИНС Unet обучается путем минимизации функции потерь, которая сравнивает выход декодера с ожидаемым сгенерированным изображением. В качестве функции потерь используется коэффициент Дайса для оценки сходства между предсказанным и исходным изображением, который вычисляется согласно (1):

$$K = \frac{2*|A \cap B|}{|A|+|B|}, \quad (1)$$

где A и B – два сравниваемых множества, $|A|$ и $|B|$ – размеры этих множеств, а $|A \cap B|$ – размер их пересечения. Коэффициент принимает значения от 0 до 1, где 1 – идеальное совпадение, 0 – полное отсутствие перекрытия [2].

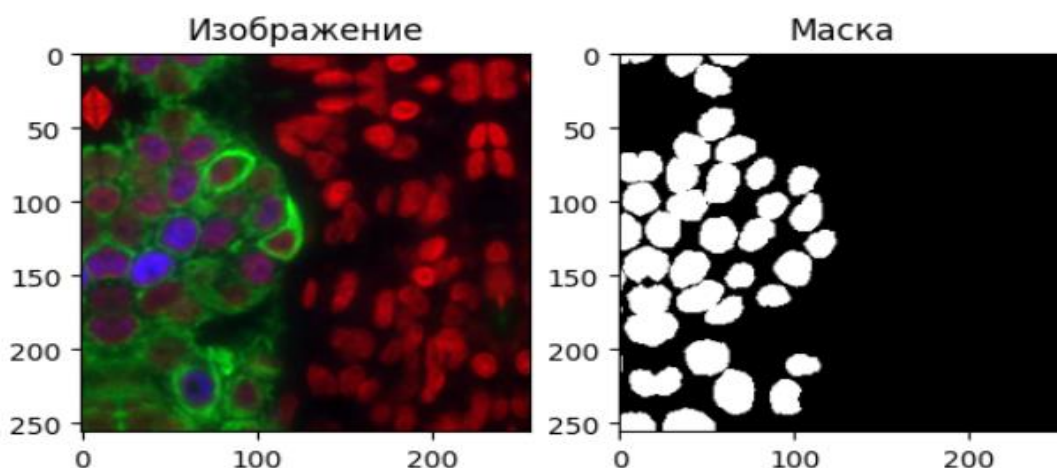


Рис. 1. Пример изображения раковых клеток и маски

Для обучения использовалось 4500 изображений, из которых 10 % использовались для валидации.

ТЕСТИРОВАНИЕ МОДЕЛИ

Для оценки полученной модели используется функция из библиотеки keras evaluate. В качестве параметров принимает тестовую выборку и возвращает результаты оценки. Результат оценки модели с помощью функции evaluate представлен в таблице.

Результат оценки модели с помощью функции evaluate

Метрика	Значение коэффициента	Потери
Коэффициент Дайса	0,8667	-0,8740

Для прогнозирования результатов сегментации использовалась функция predict. В качестве входных аргументов данная функция принимает набор изображений, для которых требуется предсказать маски сегментации раковых клеток. Выходом функции является набор изображений, соответствующих предсказанным маскам раковых клеток для входного набора изображений. Пример предсказанной маски представлен на рис. 2.

Для оценки качества сегментации использовалась формула (2), которая позволяет оценить точность контура объекта:

$$\delta = \left(1 - \frac{\sum xor(R^*, A)}{\sum or(R^*, A)}\right), \quad (2)$$

где R^* – идеальная маска объекта, A – маска объекта, полученная в результате сегментации объекта, or – дизъюнкция, xor – сложение по модулю 2 [3, с. 75]. Полученное таким образом среднее значение точности сегментации для каждого изображения из тестовой выборки составило 75,71%.

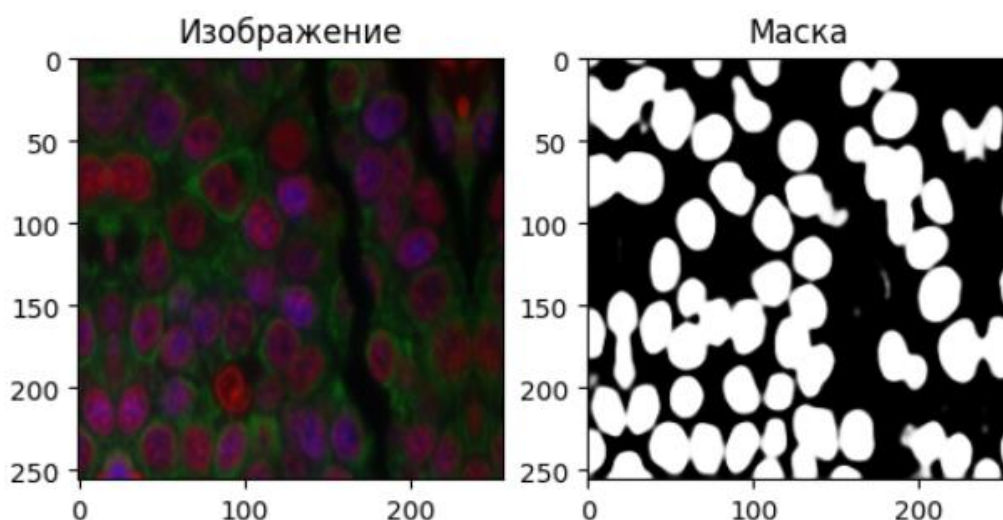


Рис. 2. Пример предсказанной маски.

ЗАКЛЮЧЕНИЕ

В работе проведена сегментация изображений раковых клеток с помощью нейронной сети UNet с использованием коэффициента Дайса в качестве метрики. Модель была обучена и протестирована на подготовленном наборе изображений. Полученная точность модели составила 75,71%.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. U-NET для решения задачи сегментации медицинских изображений Пятая Международная научно-практическая конференция «BIG DATA and Advanced Analytics. BIG DATA и анализ высокого уровня», Минск, Республика Беларусь, 13-14 мая 2019 года // редкол: О.В. Козлова, Е.Ю. Куница, М.М. Лукашевич. Минск. 2020.
2. *Петрова О.Г.* Сегментация МРТ-изображений с использованием нейронных сетей. Санкт-Петербург. 2021
3. *Лисица Е.В.* Анализ изображений популяций раковых клеток в системах обработки данных люминесцентной микроскопии: Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.01 – системный анализ, управление и обработка информации. Минск. 2021. С. 156.

СИНТЕЗ СТРУКТУРНО НЕОДНОРОДНЫХ БЛОКОВ НА ОСНОВЕ ТЕХНОЛОГИЙ ГЛУБОКОГО ОБУЧЕНИЯ

С.В. Малый, А.С. Орлова

Белорусский государственный университет, Минск, Беларусь

E-mail: maly@bsu.by

Предложена методика синтеза структуры блоков, обладающих заданными электромагнитными свойствами, на основе метода минимальных автономных блоков, методики усредненных матриц рассеяния и технологий глубокого обучения нейросетей. Приведены результаты синтеза структуры блоков, в состав которых входят кубические металлические частицы.

Ключевые слова: неоднородный блок; усредненная матрица рассеяния; метод минимальных автономных блоков; синтез структуры; технология глубокого обучения нейронных сетей.

ВВЕДЕНИЕ

Использование структурно неоднородных материалов является одним из важнейших направлений разработки и совершенствования электродинамических систем микроволнового и оптического диапазонов. Среди них следует выделить композиты, метаматериалы, метаповерхности, электромагнитные и фотонные кристаллы [1]. Для моделирования электродинамических систем, в состав которых входят неоднородные материалы, перспективным является использование метода минимальных автономных блоков [2] и методики усредненной матрицы рассеяния автономного блока [3]. Усредненная матрица рассеяния неоднородного блока полностью определяется его волновыми размерами, внутренней структурой и материальным составом. Представляет практический интерес решение задачи синтеза внутренней структуры блока по заданной матрице рассеяния с использованием технологий глубокого обучения искусственных нейронных сетей.

ТЕХНОЛОГИЯ СИНТЕЗА СТРУКТУРЫ НЕОДНОРОДНЫХ БЛОКОВ

Описание электромагнитных свойств композитов и метаматериалов усредненной матрицей рассеяния обладает большими потенциальными возможностями. Усредненная матрица рассеяния трехмерных блоков имеет 12 порядок и является альтернативой описанию неоднородных сред эффективными материальными параметрами в скалярной или тензорной форме. Важной особенностью усредненной матрицы рассеяния является ее локальный характер, т.е. она является характеристикой одиночного блока, а не усреднения по большому фрагменту материала. Задача синтеза

композитов и метаматериалов заключается в нахождении структуры пространственного распределения материалов из заданного набора с учетом схемы декомпозиции макроблоков на систему однородных блоков.

Рассмотрим методику синтеза внутренней структуры неоднородного блока с использованием технологии глубокого обучения. Для простоты рассмотрим макроблок, содержащий $5 \times 5 \times 5$ внутренних блоков. В простейшем случае в состав неоднородного макроблока могут входить два типа материалов: идеальный проводник и вакуум, причем наличие металлических блоков возможно только в центральной части макроблока с размерами $3 \times 3 \times 3$ блока.

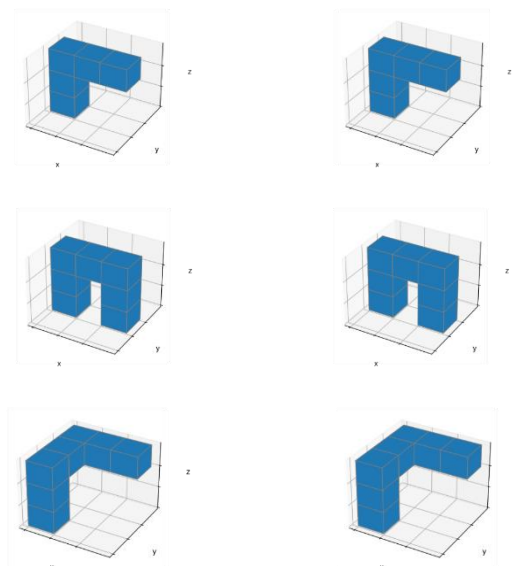
Исходной информацией для синтеза структурно неоднородного макроблока является усредненная матрица рассеяния. Синтез макроблока заключается в нахождении внутренней структуры блока, для которой усредненная матрица рассеяния наиболее близка к целевой в заданной метрике.

Для реализации процедуры синтеза использована полносвязная нейронная сеть, состоящая из входного, выходного и двух скрытых слоев, содержащих 256 и 128 нейронов соответственно. Функция активации для входного слоя – “ReLU”, для выходного слоя – “Sigmoid”. Метод оптимизации нейронной сети – “RMSprop”. Функция потерь – бинарная кросс-энтропия. Размерность входного слоя равна 288, что соответствует суммарному количеству действительных и мнимых частей элементов матрицы рассеяния. Размерность выходного слоя равна 27, что соответствует количеству блоков в варьируемой части макроблока.

Объем выборки данных, на которых проводилось обучение сети, составляло 470000 вариантов, причем 80 % этих данных использовались для обучения, а 20 % – для оценки качества обучения сети. Количество эпох обучения равно 50.

Для формирования обучающей выборки были рассчитаны усредненные матрицы рассеяния для различных вариантов распределения металлических кубиков в варьируемой области макроблока, количество которых и распределение задавалось генераторами случайных чисел. Расчет усредненной матрицы включал два этапа: определение матрицы рассеяния многоканального блока, порядок которой равен 300; подключение к каждой грани макроблока блока перехода от многоканального представления электромагнитного поля к двухмодовому. Порядок усредненной матрицы рассеяния равен 12. Внутренние блоки имели кубическую форму с размером 1 мм. Размер макроблока равен 5 мм. Расчеты усредненных матриц рассеяния проводились на длине волны 50 мм.

На рисунке представлены результаты тестирования обученной нейронной сети в виде рисунков с синтезированной структурой центральной части макроблоков (слева) и тестовой (справа). Блоки, заполненные металлом, выделены синим цветом.



Распределение металлических блоков в макроблоке

Анализ результатов демонстрирует возможность использования разработанной нейронной сети для синтеза внутренней структуры макроблоков по заданной усредненной матрице рассеяния. При этом суммарный объем вариантов в обучающей и тестовой выборке составил 0,35% от максимально возможного числа вариантов.

ЗАКЛЮЧЕНИЕ

Установлена принципиальная возможность использования технологии глубокого обучения для синтеза внутренней структуры блоков по заданной усредненной матрице рассеяния, характеризующей их электромагнитные свойства.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Engheta Nader and Ziolkowski Richard W.* Metamaterials: physics and engineering explorations. Wiley. 2006.
2. *Никольский В.В., Никольская Т.И.* Декомпозиционный подход к задачам электродинамики. М.: Наука. 1983. С. 304.
3. *Maly S.V.* Homogenization of metamaterials on the basis of average scattering matrixes// Abstracts on International conference «DAYS ON DIFFRACTION'2010». Saint Petersburg, Russia. June 8–11, 2010. P. 114.

ПРИМЕНЕНИЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ ДЛЯ ФОРМИРОВАНИЯ ОБУЧАЮЩЕГО НАБОРА ПРИ СЕГМЕНТАЦИИ ЯДЕР КЛЕТОК РАКОВЫХ ТКАНЕЙ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ

Сылунь Сьюй, В.В. Скакун

Белорусский государственный университет, Минск, Беларусь;

E-mail: xusilun@hotmail.com

Сегментация и подсчет ядер клеток на гистологических и иммуногистохимических изображениях срезов раковых тканей позволяет сделать оценку степени развития заболевания. Основным инструментом для выполнения сегментации изображений являются нейронные сети глубокого обучения. В данной работе произведено исследование применения метода главных компонент для расширения обучающего набора при сегментации ядер клеток на флуоресцентных изображениях раковых тканей молочной железы. Наилучшие результаты получены при добавлении первой главной компоненты с последующей стандартизацией данных.

Ключевые слова: глубокое обучение; изображения раковых клеток; предварительная обработка; метод главных компонент; U-Net.

ВВЕДЕНИЕ

В настоящее время цифровая обработка изображений стала важной частью медицинской диагностики и имеет широкий спектр применения. Анализ гистологических и иммуногистохимических изображений составляет основу диагностики многих видов рака. Важным шагом в процессе анализа таких изображений является сегментация ядер раковых клеток, целью которой является отнесение каждого пикселя к классу пикселей, образующему или область ядра, или фона. Искусственные нейронные сети, особенно сверточные нейронные сети (CNN), сегодня являются основным инструментом решения данной задачи.

Основной проблемой применения нейронных сетей к решению задачи сегментации является необходимость формирования сравнительно большого набора изображений, размеченных экспертами. Поскольку мы имеем ограниченный набор размеченных изображений, всего 9 изображений размером 2048×2048 пикселей, критическим является проведение предварительной подготовки данных. Нами уже было проведено исследование различных методов предварительной подготовки данных для обучения нейронной сети [1]. Наилучшим явился метод, заключающийся в выделении областей размером 256×256 пикселей с перекрытием наполовину как по вертикали, так и по горизонтали, с последующим выделением из каждой полученной области по 8 фрагментов размером 128×128 пикселей со случайным смещением. Поскольку в обучающем наборе вследствие

перекрытия имеется значительное число одинаковых фрагментов, для устранения эффекта переобучения необходимо применять аугментацию изображений.

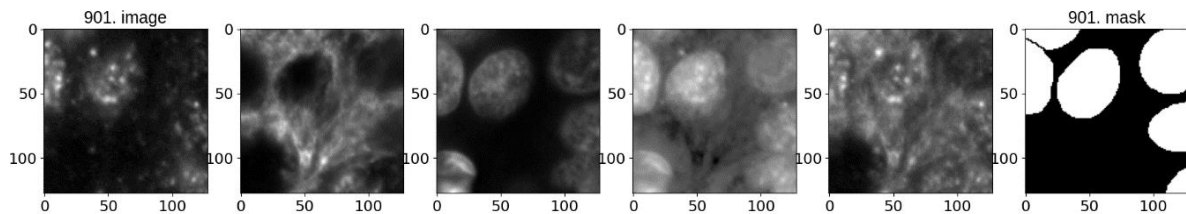
Исследования показали, что двухкратное увеличение полученной обучающей выборки путем повторной подачи изображений на вход нейронной сети (с выполнением аугментации) позволяет улучшить качество сегментации (с средним на 1-2% по метрике IoU) и не переобучить модель. Следовательно, подготовленный набор данных еще не является оптимальным. Одним из способов увеличения объема полезной информации является применение различного рода преобразований исходных данных. Если исходными данными являются изображения, то такими преобразованиями являются аффинные преобразования, изменение яркости, контрастности и т.д., входящие в набор методов аугментации изображений. К таким преобразованиям также можно отнести методы снижения размерности, в частности метод главных компонент (МГК). Применение МГК позволило значительно улучшить качество сегментации ядер раковых клеток на флуоресцентных изображениях с использованием классических методов сегментации изображений [2]. Так как мы работаем с трехканальными флуоресцентными изображениями раковых клеток, где вследствие перекрытия спектров флуорофоров имеется корреляция данных в соседних каналах, то добавление еще одного канала, являющегося первой главной компонентой, построенной на основе трех исходных каналов, может предоставить дополнительную информацию для модели и улучшить качество сегментации. Соответственно, целью работы является исследование применения МГК для расширения обучающего набора при сегментации ядер клеток раковых тканей с помощью нейронных сетей глубокого обучения.

РЕЗУЛЬТАТЫ

Исходными данными являются 9 флуоресцентных изображений срезов ткани опухоли молочной железы в формате RGB, полученные при 10-кратном увеличении с использованием эпифлуоресцентного инвертированного микроскопа Nikon TE200, оснащенного ПЗС-камерой Photometrics серии 300. Изображения 3-х канальные, 2048x2048 пикселей с разрешением 0,2 мкм/пиксель [2]. Синий краситель 4,6-диамино-2-фенилиндола дигидрохлорид (DAPI) использовался для маркировки всех ядер. Красный цианиновый краситель Cy5 использовался для маркировки белка рецептора эстрогена, содержащегося в основном в ядрах раковых клеток. Белок цитокератин, содержащийся в цитоплазме раковых клеток, был помечен зеленым цианиновым красителем Cy3. Таким образом, маркерами раковых клеток являются два красителя, Cy5 и Cy3. Размеченные

экспертами изображения, на которых выделены только маски ядер клеток (ground truth, GT), представляют собой бинарные изображения, в которых пикселям ядер клеток присвоено значение 1.

В процессе предварительной подготовки данных, заключающейся в выделении областей размером 256×256 пикселей с перекрытием наполовину как по вертикали, так и по горизонтали с последующим выделением из каждой полученной области по 8 фрагментов размером 128×128 пикселей со случайным смещением, получено 12064 фрагмента размером 128×128 . Для каждого фрагмента независимо был применен МГК. Признаками в преобразовании являются каналы изображения и результатом применения МГК являются изображения той же размерности, что и исходное. Перед применением МГК была произведена стандартизация данных (приведение к нулевому среднему и единичной дисперсии). Так как мы имеем всего 3 признака и дисперсия, приходящаяся на главные компоненты, быстро падает, информативными являются только первые две компоненты. На рисунке показан результат применения МГК для фрагмента изображения раковых клеток размером 128×128 пикселей.



Фрагмент иммуногистологического флуоресцентного изображения раковых клеток размером 128×128 пикселей. Слева направо: красный канал, зеленый канал, синий канал, первая и вторая главные компоненты, размеченные маски ядер.

Таким образом, исходными данными для обучения являются уже 4 каналные изображения, где 4-ый канал есть 1 главная компонента, вычисленная на основе каналов исходных изображений. Были рассмотрены 4 варианта: 1) исходные данные (R,G,B); 2) исходные данные и первая главная компонента (R,G,B,PC1(RGB)); 3) исходные данные плюс первая главная компонента, вычисленная на основе нелинейных комбинаций компонент (G,B,PC1(G,B,G/B,B/G,1/G,1/B)), показавшая наилучшие результаты в работе [2]; 4) первые две главные компоненты (PC1(RGB), PC2(RGB)). Во всех вариантах проводилась стандартизация исходных данных (приведение к нулевому среднему и единичной дисперсии). Поскольку дисперсия первой главной компоненты более чем в 2 раза превышала дисперсию исходных стандартизированных данных, то дополнительно были исследованы варианты выравнивания дисперсий по всем каналам, исключая первую главную компоненту (обозначим их как 2* и 3*).

Для исследования применения МГК для расширения обучающего набора данных использовалась модель U-Net. Архитектура сети и ее параметры подробно описаны в работе [1]. После каждого слоя свертки в кодере и декодере дополнительно был добавлен слой Batch normalization. Все тесты проводились на компьютере, оснащённом графической картой RTX-3060 с объемом оперативной памяти 12 ГБ. Программная реализация выполнена на языке Python с использованием 2 версии фреймворка TensorFlow. Для всех рассматриваемых случаев размер выборки BatchSize был установлен в 32, количество итераций – 30. Для оценки качества сегментации выполнялся расчет двух метрик для тестовых данных: бинарное пересечение над объединением на уровне 0,5 (BIOU) и функция бинарной энтропии loss. Результаты исследования на основе 9 независимых экспериментов представлены в таблице. Наилучшие результаты были получены для вариантов 2 и 3. Наихудшим явился вариант с использованием только преобразованных данных. Добавление главной компоненты позволило значительно улучшить стабильность процесса обучения.

Средние значение и стандартные отклонения метрик BIOU и loss

Вариант	1	2	2*	3	3*	4
BIOU	0,876±0,003	0,878±0,001	0,877±0,003	0,878±0,001	0,877±0,004	0,845±0,005
Loss	0,153±0,007	0,151±0,003	0,154±0,005	0,150±0,001	0,153±0,008	0,197±0,008

ВЫВОД

Проведено исследование применения МГК для расширения обучающего набора при сегментации ядер раковых клеток с помощью нейронных сетей глубокого обучения. Наилучший результат был получен для обоих вариантов с добавлением первой главной компоненты и стандартизацией исходных данных (результаты сравнимы), что подтверждает идею о целесообразности применения МГК для расширения обучающей выборки.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Xu, S., Skakun, V. Comparison of Deep Learning Preprocessing Algorithms of Nuclei Segmentation on Fluorescence Immunohistology Images of Cancer Cells // PRIP 2021. Communications in Computer and Information Science, 2022, vol 1562. Springer, Cham. P. 166–177.
2. Lisitsa Y.V., Yatskou M.M., Apanasovich V.V., Apanasovich T.V. Algorithm for automatic segmentation of nuclear boundaries in cancer cells in three-channel luminescent images. // Journal of Applied Spectroscopy. 2015. 82(4) C. 598-607.

РАСПОЗНАВАНИЕ И АНАЛИЗ МОРСКИХ СУДОВ МЕТОДАМИ ГЛУБОКОГО ОБУЧЕНИЯ НА БАЗЕ АЭРОКОСМИЧЕСКИХ ДАННЫХ

И.В. Саечников, В.В. Скакун, Э.А. Чернявская

*Белорусский Государственный Университет, Минск, Беларусь
E-mail: saetchnikovivan@gmail.com*

В работе рассматривается применение методов глубокого обучения для обработки аэрокосмических данных с целью распознавания и анализа морских судов. Продемонстрированы преимущества использования комбинированного метода для решения существующих проблем детектирования подобных объектов, и достигнута точность распознавания на уровне 0,8491.

Ключевые слова: распознавание образов; детектирование объектов; глубокое обучение; дистанционное зондирование; обратная задача.

ВВЕДЕНИЕ

Задача распознавания морских судов играет ключевую роль в эффективном морском мониторинге [1], управлении логистическими цепочками и портами [2] и обеспечении безопасного судоходства [3]. Только торговый флот насчитывает более 100 тысяч зарегистрированных морских судов [4]. Хотя автоматическая идентификационная система позволяет отслеживать суда, периодически она страдает как от дороговизны коммерческих данных, так и ошибок или пропусков данных [5]. С другой стороны, согласно прогнозам по результатам анализа преимущественно публичных данных частных-государственных проектов [6-7], до 2032 года ожидается рост рынка сервисов анализа аэрокосмических данных (14,8%). Ключевую роль в создании автоматических систем компьютерного зрения для распознавания и отслеживания морских судов могут сыграть адаптивные методы глубокого обучения, которые уже доказали свою эффективность в смежных задачах распознавания и отслеживания дорожных объектов по данным дистанционного зондирования (ДЗ) [8-10]. Задача распознавания и отслеживания морских судов менее исследована на текущий момент. Таким образом, учитывая преимущество охвата больших пространственных и временных масштабов, спутниковое ДЗ и методы глубокого обучения представляют собой эффективные и экономичные инструменты для решения задачи мониторинга морских судов.

Текущие исследования по разработке и распознаванию морских судов методами глубокого обучения можно разделить на два класса – двухэтапные и одноэтапные методы. В рамках первого класса на базе архитектуры R-CNN, авторы статьи [11] представили улучшенную версию Faster-RCNN для обнаружения морских судов, используя улучшенную функцию ROI для извлечения признаков и улучшенный классификатор на базе

SVM. Среди второго класса методов в 2023 году в статье [12] предложен усовершенствованный алгоритм на базе архитектуры YOLOv7 с адаптацией anchor box для лучшей передачи характеристик судов разного размера. Несмотря на существенный прогресс в данной области, некоторые аспекты до сих пор требуют внимания, например, пропуски морских судов малых размеров, неверная идентификация рядом находящихся судов.

РЕЗУЛЬТАТЫ

Исследуемые объекты рассматривались на аннотированном наборе данных Airbus SPOT Satellite Imagery [13], содержащих более 200 тысяч изображений с разрешением 1,5 м с наблюдаемой несбалансированностью исследуемых классов. Суда на изображениях различались по размерам и находились как в открытом море, так и в доках, пристанях и т.д. — следовательно, требовалось учитывать разнообразие фонов при формировании карты признаков. Учитывая данный аспект, был предложен комбинированный метод: на первом этапе наличие судов определялось классификатором DenseNet с 121 слоем и кросс-валидацией, а на втором этапе объекты на изображениях с потенциальными морскими судами распознавались и локализовались с использованием архитектуры YOLOv8mod. Таким образом, классификатор боролся с фильтрацией бликов волн, ярких пятен, элементов облаков, которые могут определяться как морские суда ложными срабатываниями и ухудшать итоговую точность распознавания. Для фильтрации фона и упрощения распознавания сгруппированных морских судов с размытыми границами на общем фоне вместо стандартных ограничительных рамок использовались ориентированные ограничительные рамки. Для предотвращения слияния объектов были сегментированы области границ каждого объекта и области их соприкосновения. Перед обучением на тренировочном наборе данных была применена вертикальная, горизонтальная и диагональная флип-аугментация, а также аугментация поворота. Для эффективного использования GPU мы выделяли из исходных изображений фрагменты размером 512x512, центрированные на отмеченные объекты кораблей. Точность распознавания судов оценивалась по усреднению по набору порогов F2 score. Таким образом, на тестируемом наборе предложенным комбинированным методом F2 достигнута точность 0,8491.

ЗАКЛЮЧЕНИЕ

В данной работе предложен комбинированный метод распознавания и анализа морских судов на основе методов глубокого обучения. Стоит отметить, что предложенный метод распознавания и анализа морских судов является перспективным в решении ряда обратных задач. Эти задачи

связаны не только с реконструкцией и анализом объектов по семантическим признакам, но также применим в аналитических задачах смежных областей, таких как управление рисками в логистической отрасли, дистанционное зондирование и одной из самых актуальных быстрорастущих областей – spatial finance.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Salerno, E.; Di Paola, C.; Lo Duca, A. Remote Sensing for Maritime Monitoring and Vessel Identification // Remote Sensing 2024. P. 16, 776. DOI: 10.3390/rs16050776.
2. Worlds of Access or Absence: Supply Security and Maritime Security in an Era of Intense Geopolitical Competition [Электронный ресурс]. URL: <https://hcss.nl/report/worlds-of-access-or-absence-supply-security-and-maritime-security> (дата обращения: 19.03.2024).
3. Acciaro, Michele & Serra, Patrizia. Maritime Supply Chain Security: A Critical Review. // IFSPA: Trade, Supply Chain Activities and Transport: Contemporary Logistics and Maritime Issues. 2013.
4. Merchant Fleet Infographic: 2023 Update [Электронный ресурс]. URL: <https://www.ener8.com/> (дата обращения: 25.03.2024).
5. Pohontu, Alexandru. A Review over AI Methods Developed for Maritime Awareness Systems. // Scientific Bulletin of Naval Academy. XXIII. 2020. P. 287-299. DOI: 10.21279/1454-864X-20-I2-107.
6. Satellite Imaging Market - By Application by End-User & Forecast, 2024 – 2032 [Электронный ресурс]. URL: <https://www.gminsights.com/industry-analysis/satellite-imaging-market> (дата обращения: 20.03.2024).
7. B. Caldecott, M. McCarten, C. Christiaen, C. Hickey. “Spatial finance: practical and theoretical contributions to financial analysis”, Journal of Sustainable Finance & Investment. 2022.
8. Deep Neural Network-Based Dynamical Object Recognition and Robust Multiobject Tracking Technique for Onboard Unmanned Aerial Vehicle’s Computer Vision-Based Systems / Saetchnikov I. [et al.] // IEEE Journal on Miniaturization for Air and Space Systems, vol. 4, no. 3. 2023. P. 250-256. DOI: 10.1109/JMASS.2023.3274929.
9. biLSCCS: modular dynamical on-road objects trajectory prediction approach / Saetchnikov I. [et al.] // IEEE 10th International Workshop on Metrology for AeroSpace. 2023. P. 131-135. DOI 10.1109/MetroAeroSpace57412.2023.10190032.
10. Efficient objects tracking from an unmanned aerial vehicle / Saetchnikov I. [et al.] // IEEE 8th International Workshop on Metrology for AeroSpace (MetroAeroSpace). 2021. P. 221-225. DOI: 10.1109/MetroAeroSpace51421.2021.9511748.
11. MS-SSD: Multi-scale single shot detector for ship detection in remote sensing images. / Wen, G. [et al.] // Applied Intelligence. 2023. 53(2), P. 1586–1604. DOI: 10.1007/s10489-022-03549-6.
12. A lightweight YOLOv5 optimization of coordinate attention. / Wu, J. [et al.] // Applied Sciences. 2023. 13(3) P. 1746. DOI: 10.3390/app13031746.
13. 1.5m Satellite Imagery | SPOT [Электронный ресурс]. URL: <https://intelligence.airbus.com/imagery/our-optical-and-radar-satellite-imagery/spot/> (дата обращения: 15.03.2024).

ПРОБЛЕМА КАЧЕСТВА ДАННЫХ ДЛЯ МАШИННОГО ОБУЧЕНИЯ

А.А. Халин

ФГБОУ ВО «Курский государственный университет», Курск,
Российская Федерация

E-mail: khalin_andrey@rambler.ru

В работе проводится обсуждение проблемы качества наборов данных, необходимых для обучения или дообучения моделей машинного обучения. Рассматриваются способы намеренного изменения данных с целью воздействия на модель машинного обучения заранее заданным образом. Приводятся способы защиты от проводимых атак.

Ключевые слова: машинное обучение; наборы данных; качество данных; атака отравления данных.

В последние годы продолжает наблюдаться высокий интерес как к искусственному интеллекту (ИИ) в целом, так и к машинному обучению, являющемуся одним из направлений в ИИ, в частности. Это подтверждается, например, ежегодными отчетами, выпускаемыми Стэнфордским университетом [1], анализами публикационной активности [2] и т.д.

При этом важным вопросом при развитии методов машинного обучения продолжает оставаться подбор качественных наборов данных («набор данных (dataset): совокупность данных, в том числе соответствующих им метаданных, организованных по определенным правилам и принципам описания» [3]), на которых будет производиться обучение модели. Например, в [4] качество данных (наборов данных) определяется, как понимание или оценка их пригодности для достижения цели, поставленной в рамках предметной области. В [5] же под качеством понимается «степень соответствия совокупности присущих характеристик объекта требованиям». Там же можно найти, что «данные – факты об объекте». Можно привести еще ряд качества данных, но все они в общих чертах будут схожи.

Проблема качества данных появилась не вчера. Во все времена от этого зависело очень много. Можно привести большое количество примеров из истории, но данная работа не целена на это, поэтому не будем здесь на этом останавливаться на этом. Отметим лишь, что попытки разработки «широкой» концепция качества данных предпринимаются не так давно. Здесь в качестве примера можно привести, например, работу [6], в которой авторы выявили общие закономерности, связанные с качеством данных, а также привели рекомендации для специалистов по информационной безопасности.

Несмотря на сказанное выше, именно сейчас проблема качества данных стоит особо остро. В том же отчете [1], говорится, что медицина является областью, в которой искусственный интеллект и машинное обучение используются наиболее активно. При этом растет количество дезинформации и подделок. И если речь идет о жизни и здоровье людей, то нельзя считаться с такими подделками, какими бы они ни были. Поэтому перейдем теперь к обсуждению вопроса о намеренном изменении данных.

В проводимых исследованиях мы можем использовать различные модели машинного обучения: разработанные непосредственно нами или же разработанные и предобученные другими специалистами, но в любом случае при работе с этими моделями мы будем использовать некоторые данные. Чаще всего эти данные берутся из открытых источников, например, распространенным хранилищем различных наборов данных является система Kaggle. При этом у нас нет никакой уверенности в качестве этих данных, которые могут быть модифицированы (испорчены) как умышленно, так и неумышленно. Здесь стоит отметить, что нами никак не рассматривается тот факт, что воздействие может происходить непосредственно на модель машинного обучения (МО) с целью ее реакции заданным образом. Также мы не касаемся того факта, что предобученная модель может давать некорректные результаты, так как предобучение проходило на испорченных данных. И получение некорректных результатов будет сохраняться в такой модели даже если мы потом используем качественные данные. Если мы говорим о неумышленной «порче» данных, то она может быть связана с ошибками, допущенными при проведении разметки, и связана во многом с уровнем подготовки специалистов, занимающихся разметкой.

Говоря же о преднамеренной модификации данных, можно выделить следующие способы такого воздействия [7-8]:

- Изменение разметки данных;
- р-фальсификация;
- Использование чистых меток (атака столкновением признаков, атаки уклонением и др. [9 - 10]);
- Непосредственная подготовка некачественных данных и т.д.

При наличии такого числа различных способов проведения атак на модели МО естественным является вопрос противостояния атакам. Выделяют несколько способов противостояния атакам на модели МО [11]:

- Методы маскировки градиента;
- Оборонительная дистилляция;
- Состязательное обучение;
- Реконструкция входных данных и т.д.

При этом нужно отметить, что несмотря на наличие способ противостояния нужно быть аккуратными при их использовании, так как имеются способы их обхода, например, в работах [12 - 13] говорится о таких обходах. Таким образом, как бы нам не хотелось положиться на те или иные методы защиты, необходимо тщательно подбирать набор данных и/или модель МО для работы, чтобы не получить результат, который принесет дополнительные проблемы в последующих исследованиях или при внедрении их, например, в здравоохранение, образование и т.д.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. 2023 AI Index Report [Электронный ресурс]. URL: <https://aiindex.stanford.edu/report/> (дата обращения: 20.13.2024).
2. Hao, K. We analyzed 16,625 papers to figure out where AI is headed next // MIT Technol. Rev. 2019. С. 1-19.
3. ГОСТ Р 59898– 2021. Национальный стандарт российской федерации. ОЦЕНКА КАЧЕСТВА СИСТЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: национальный стандарт Российской Федерации: издание официальное: утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 26 ноября 2021 г. 1620-ст. Москва: Российский институт станд, 2021. С. 24.
4. *Rupa Mahanti*. Data Quality: Dimensions, Measurement, Strategy, Management, and Governance. Quality Press. 2019.
5. ISO 9000:2015(ru) Системы менеджмента качества. Основные положения и словарь [Электронный ресурс]. URL: <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:ru> (дата обращения 20.13.2024).
6. Strong D. M., Lee Y. W., Wang R. Y. Data quality in context //Communications of the ACM. Т. 40. №. 5. 1997. С. 103-110.
7. *Намиот Д. Е.* Введение в атаки отравлением на модели машинного обучения //International Journal of Open Information Technologies. Т. 11. №. 3. 2023. С. 58-68.
8. *Anley C.* Practical attacks on machine learning systems. 2022.
9. *Костюмов В. В.* Обзор и систематизация атак уклонением на модели компьютерного зрения //International Journal of Open Information Technologies. Т. 10. №. 10. 2022. С. 11-20.
10. *Пришлецов Д. Е., Пришлецов С. Е., Намиот Д. Е.* Камуфляж как состязательные атаки на модели машинного обучения //International Journal of Open Information Technologies. Т. 11. №. 9. 2023. С. 41-49.
11. Атаки и методы защиты в системах машинного обучения: анализ современных исследований / Котенко И. В. [и др.] //Вопросы кибербезопасности. №. 1. 2024. С. 59.
12. *Athalye A., Carlini N., Wagner D.* Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples //International conference on machine learning. PMLR. 2018. С. 274-283.
13. *Carlini N., Wagner D.* Towards evaluating the robustness of neural networks //2017 IEEE symposium on security and privacy (sp). IEEE. 2017. С. 39-57.

СЕКЦИЯ «ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ И СИСТЕМЫ»

COMPARATIVE STUDY OF STC AND LSB IMAGE CODING METHODS BASED ON STEGANOGRAPHIC MODIFICATION STATISTICS

Shifeng Qiu, V.S. Sadv

*Belarusian State University, Minsk, Belarus
E-mail: qiushifeng2012@126.com; sadv@bsu.by*

In this paper, steganography algorithms based on STC and LSB methods are subjected to comparative analysis from different perspectives, and visualization and quantitative statistics of their results are performed.

Keywords: algorithm; low bit-plane image; STC and LSB steganographic methods; statistics

СРАВНИТЕЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДОВ STC И LSB КОДИРОВАНИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ СТАТИСТИКИ СТЕГANOГРАФИЧЕСКИХ МОДИФИКАЦИИ

Цю Шифэн, Садв В.С.

*Белорусский государственный университет, Минск, Беларусь
E-mail: qiushifeng2012@126.com; sadv@bsu.by*

В работе алгоритмы стеганографии на основе STC и LSB методов подвергаются сравнительному анализу с разных точек зрения, а также выполняются визуализация и количественная статистика результатов их использования.

Ключевые слова: алгоритм; младшая битовая плоскость изображения; STC и LSB стеганографические методы; статистика

INTRODUCTION

Information hiding is one of the important means of information security. In the process of network communication and data storage, it can not only ensure the security of the confidential information itself, but also hide the fact of transmission and storage of secret information.

Traditional, widespread, steganographic algorithms based on LSB method have a significant disadvantage - low steganographic stability.

Steganographic algorithm based on the STC method is an image encoding algorithm that minimizes the distortion of the container image when embedding secret information into it by minimizing the number of modifiable bits of the container. The mathematical basis of STC includes the following aspects:

binary linear convolution code, Syndrome-Trellis code, Trellis diagram and Viterbi algorithm [1, 2].

EXPERIMENTAL STUDIES

In this paper, experimental studies of steganographic parameters of STC and LSB methods for hiding secret information (message) are carried out.

For this purpose, five containers of two types are selected each: synthetic images and natural photos. The sizes of the containers are 2500/1500/1000/500/500/250 KB respectively. The lowest (zero) bit plane of the container was used to embed the message. The corresponding maximum possible message size is the container size divided by eight. The parameter diff % was evaluated, showing the percentage of the number of modified bits of the container to the total number of bits of the embedded message.

The container bit modification statistics when using STC and LSB methods are shown in Fig. 1. It can be seen that when the low-order bit plane of the container is filled with the message at 100%, the variation of the diff parameter did not exceed a few percent for a 10-fold change in the container size. However, the diff value is on average 5% smaller when using the STC method.

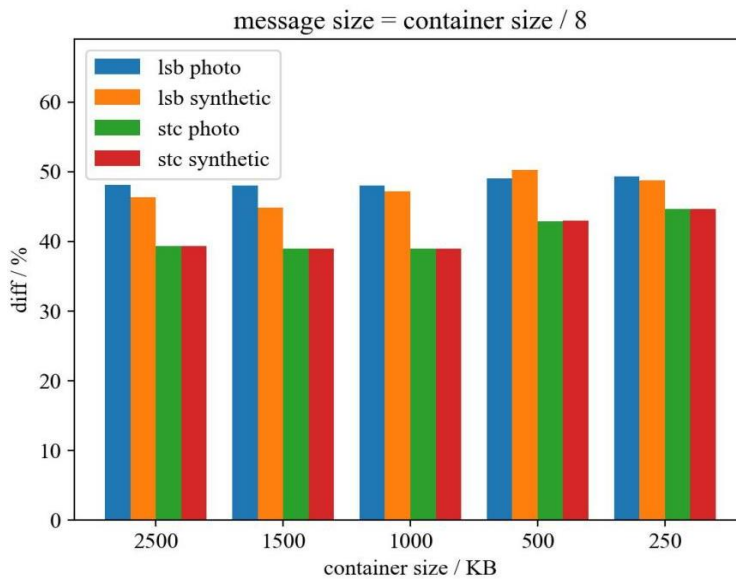


Fig. 1. Statistics of image modification with STC or LSB

Fig. 2 visualizes the results of using two types of containers as well as LSB and STC embedding methods. The message is embedded in the minor bit planes of the containers when they are 100% filled. The results of steganographic modifications of two synthesized (artificial) containers are shown on the left, and two natural (photo) containers are shown on the right. Both types of

containers have the same size - 2500 KB and 1500 KB, the message size is 300 KB and 180 KB respectively, which corresponds to the size of the minor bit planes of the containers.

It can be seen that steganofiles obtained by both methods are visually indistinguishable from unfilled containers.

At the same time, differences in the image structure of the minor bit planes of both unfilled containers and LSB STC steganofiles are observed, which can be used in steganoanalysis.

The estimate of container distortion in this experiment by PSNR parameter for both steganographic methods was about 52 dB.

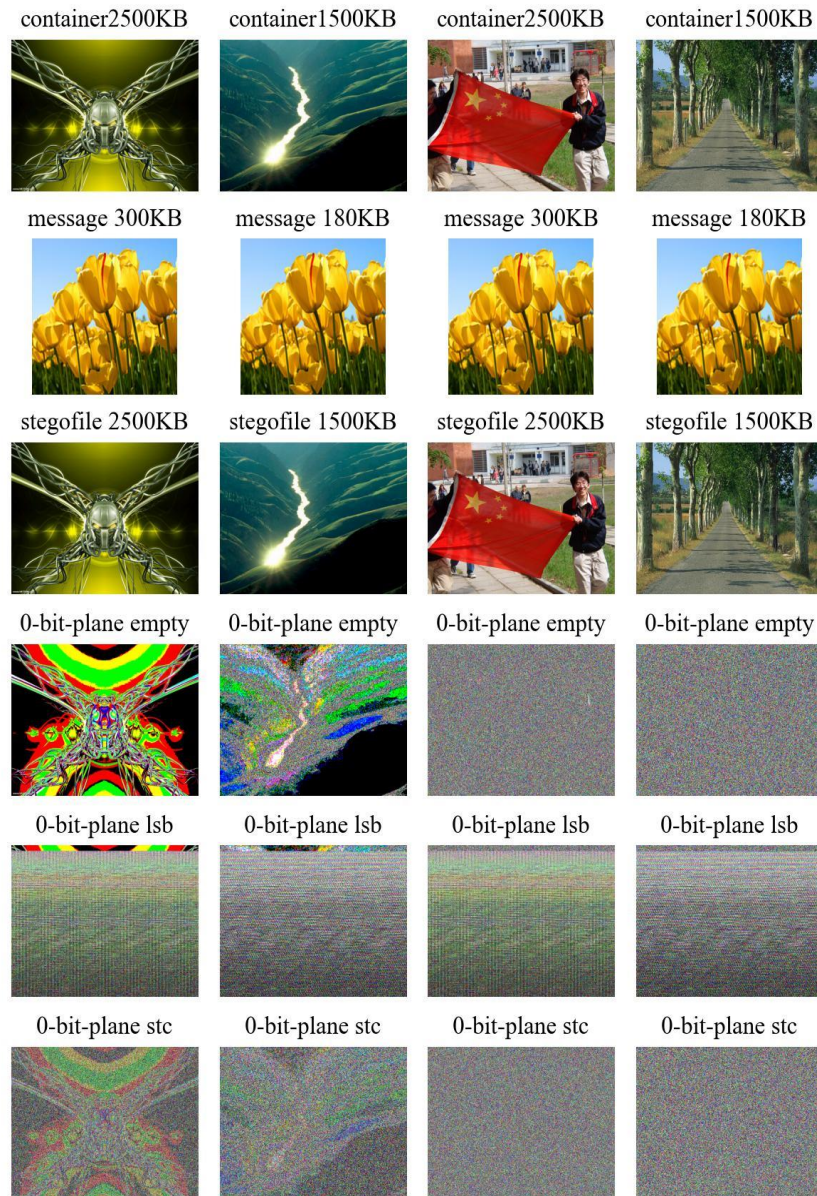


Fig. 2. Comparison of images and their 0-bit plane using STC or LSB

Fig. 3 shows the dependence of diff value on the degree of filling the container with message when using LSB and STC steganographic methods.

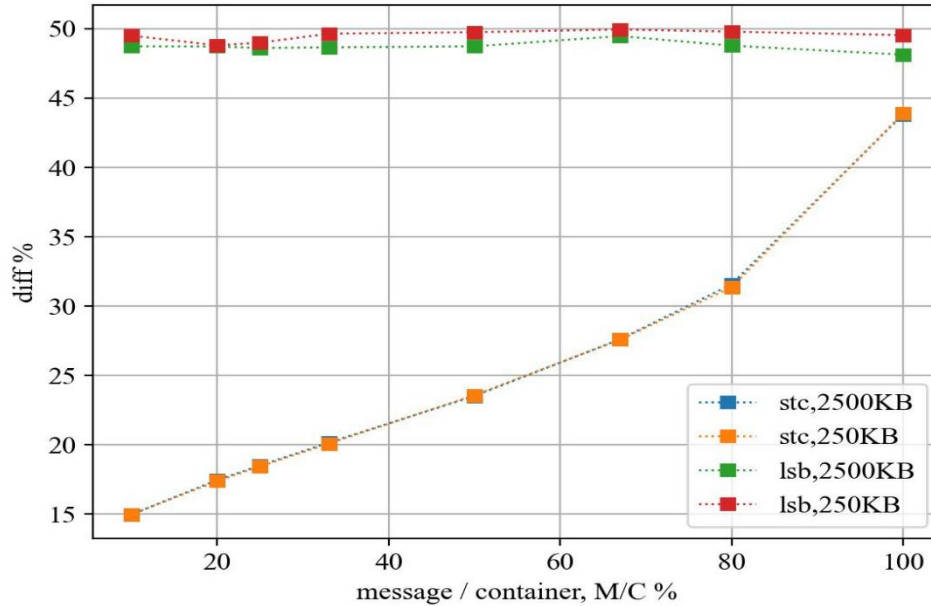


Fig. 3. Statistics of image modification using STC with different message sizes

The figure shows that the STC method is characterized by a decrease in the number of modifiable bits of the container as the size of the embedded message decreases. It is noticeable that when the container occupancy with the message decreases from 100% to 50%, the diff value decreases from 43% to 23%. The value of diff in case of LSB method remains practically unchanged at the level of 50%, which is statistically justified.

CONCLUSION

The STC steganographic method has significant advantages over the LSB method in terms of container distortion when the container is 80 percent or less filled with the message, regardless of the absolute size of the container.

REFERENCES

1. Filler T., Judas J., Fridrich J. Minimizing Embedding Impact in Steganography using Trellis-Coded Quantization // Proc. SPIE, Electronic Imaging, Media Forensics and Security XII, San Jose, CA. January 18-20, 2010.
2. Tomas Filler, Jan Judas, Jessica Fridrich. Minimizing Additive Distortion in Steganography Using Syndrome-Trellis Codes". IEEE TRANSACTIONS ON INFORMATION FOREN

СИСТЕМА ОРИЕНТАЦИИ ИНЖЕНЕРНОЙ МОДЕЛИ НАНОСПУТНИКА ПО СОЛНЦУ

Т. Е. Беляев, М. А. Павлышко, В. С. Баранова, А. О. Николаев

Белорусский государственный университет, Минск, Беларусь
E-mail: rct.belyaev@bsu.by, paulyshka@bsu.by, rct.baranovaVS@bsu.by,
rct.nikolaevAO@bsu.by

Предложен вариант построения системы определения ориентации инженерной модели сверхмалого космического аппарата по солнцу. Описаны программно-аппаратные средства для создания модуля системы определения ориентации и его функциональные возможности. Описана работа с датчиком освещенности Rohm BH1750.

Ключевые слова: наноспутник; микроконтроллеры; система солнечной ориентации.

ВВЕДЕНИЕ

Определение положения наноспутника критически важно для его корректной работы. Солнечные панели преобразуют солнечный свет в электрическую энергию, необходимую для работы спутника. Ориентация спутника относительно Солнца важна для максимизации эффективности панелей. Точное положение антенн на спутнике обеспечивает качественную связь с Землей или другими спутниками. Оптимальная ориентация зависит от расстояния до объекта связи, направления антенны и возможных помех. Положение Солнца также влияет на оптимальную ориентацию при коммуникации.

Датчик солнечного излучения — это общий термин, охватывающий класс устройств, чувствительных к солнечному излучению, используемых для определения ориентации, управления ориентацией, определения скорости вращения и определения времени [1]. В инженерной модели сверхмалого космического аппарата в качестве датчиков применяется Rohm BH1750.

Датчик освещенности Rohm BH1750 — это цифровой датчик для измерения уровня освещенности окружающей среды. Датчик состоит из фотодиода, операционного усилителя, АЦП и логической схемы, которая осуществляет конверсию данных АЦП и их передачу посредством интерфейса I2C. Диапазон измерения ограничен сверху 65535 Люкс, что позволяет использовать его в помещении и в пасмурную погоду.

ОПРЕДЕЛЕНИЕ ВЕКТОРА НАПРАВЛЕНИЯ НА СОЛНЦЕ

Освещенность датчика BH1750 является функцией угла падения θ :

$$I = I_0 \cos(\theta), \quad (1)$$

где I_0 – полная интенсивность излучения солнца, θ – угол между вектором распространения солнечных лучей и нормалью к датчику освещенности.

Предположим, что датчики освещенности, установленные на шести поверхностях наноспутника, идентичны и ортогональны друг другу. Тогда солнечный вектор может быть записан как [2]:

$$\vec{S} = (x, y, z). \quad (2)$$

Для для определения положения Солнца измеряют интенсивности лучей света, падающих на шесть граней спутника $I_{x-}, I_{x+}, I_{y-}, I_{y+}, I_{z-}, I_{z+}$ [3].

Если выбрать интенсивности I_a, I_b, I_c :

$$\begin{cases} I_a = \max(I_{-x}, I_{+x}) \\ I_b = \max(I_{-y}, I_{+y}), \\ I_c = \max(I_{-z}, I_{+z}) \end{cases} \quad (3)$$

тогда солнечный вектор можно представить в виде:

$$S = \begin{bmatrix} \frac{\pm I_a}{\sqrt{I_a^2 + I_b^2 + I_c^2}} \\ \frac{\pm I_b}{\sqrt{I_a^2 + I_b^2 + I_c^2}} \\ \frac{\pm I_c}{\sqrt{I_a^2 + I_b^2 + I_c^2}} \end{bmatrix}. \quad (4)$$

ПОЛУЧЕНИЕ ДАННЫХ

Структурная схема устройства для определения вектора направления на Солнце (формула 4) приведена на рис. 1.

На каждой боковой грани наноспутника находится плата датчика освещенности (рис. 2), состоящая из сенсора BH1750 и микроконтроллера STM32F030F4P6. При попадании солнечных лучей на датчик BH1750 происходит генерация ЭДС на фотодиоде и ее оцифровка посредством 16-битного АЦП датчика. Микроконтроллер STM32F030F4P6 формирует данные датчика с помощью программной реализации интерфейса I2C и передает их в I2C-шину СМКА посредством аппаратной реализации интерфейса I2C.

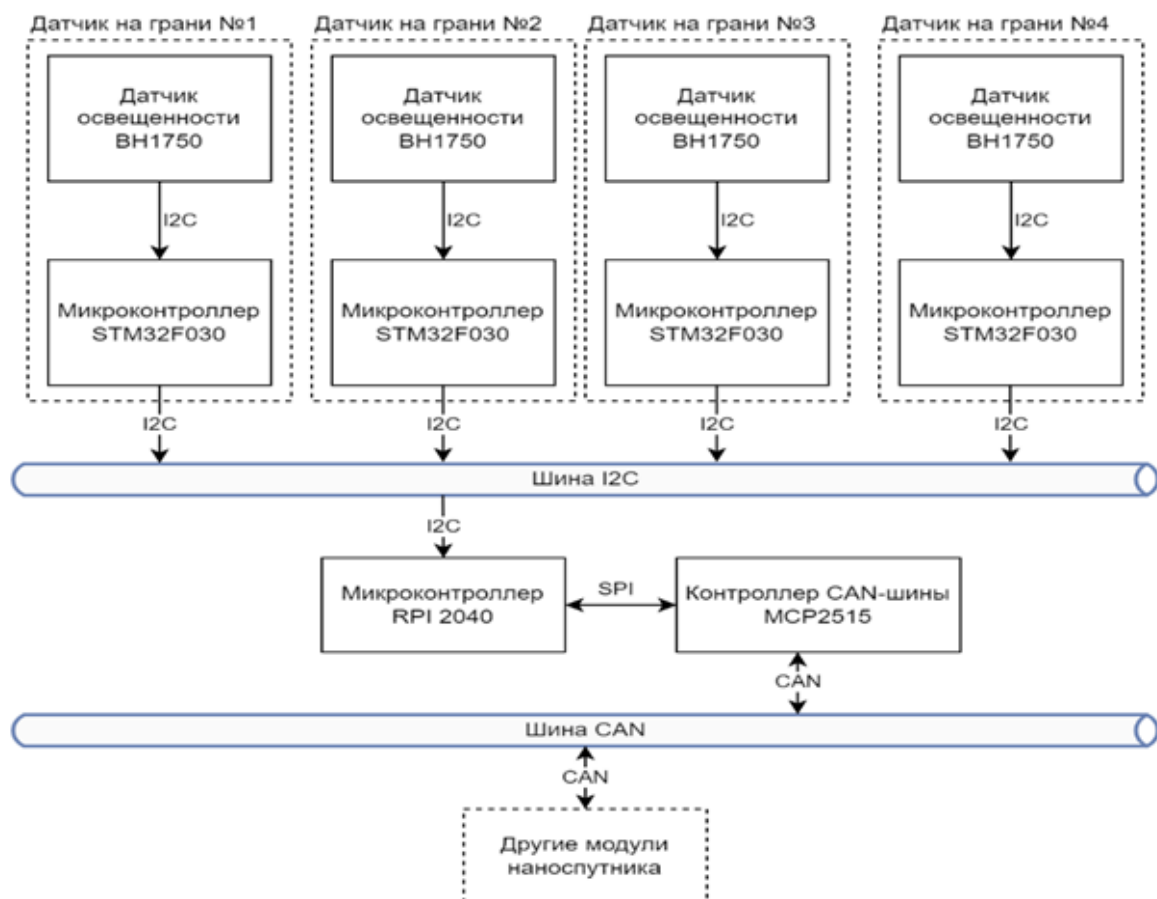


Рис. 1. Схема подключения датчиков освещенности

Микроконтроллер RPi RP2040 выполняет циклический опрос датчиков на шине интерфейса и преобразование полученного набора данных в солнечный вектор.

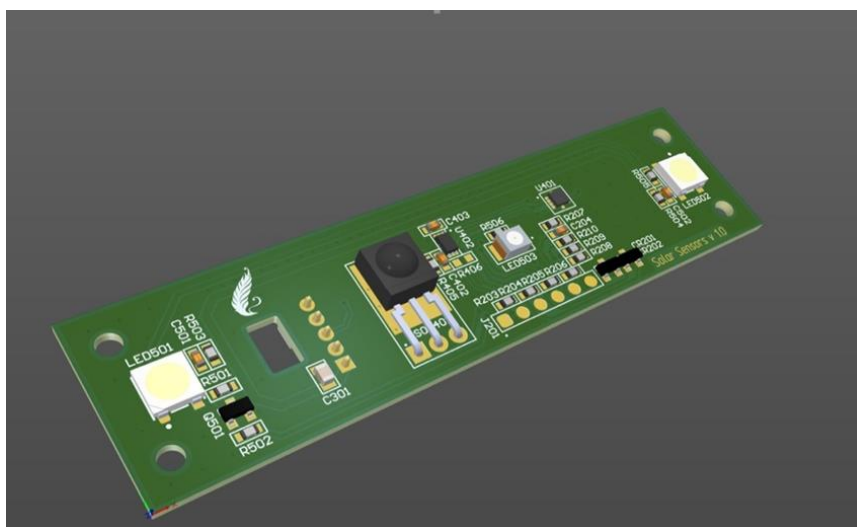


Рис. 2. Макетная плата модуля освещенности

Дальнейшие расчеты вектора направления на Солнце на основе значений датчика освещенности (формула 4) проводятся в составе комплексных алгоритмов определения ориентации наноспутника соответствующей бортовой системой.

На рис. 3, приведена зависимость интенсивности солнечного света от времени суток, полученная с помощью датчика ВН1750, который был установлен на оконном стекле. Рисунок подтверждает адекватность и реалистичность данных, полученных от датчика.

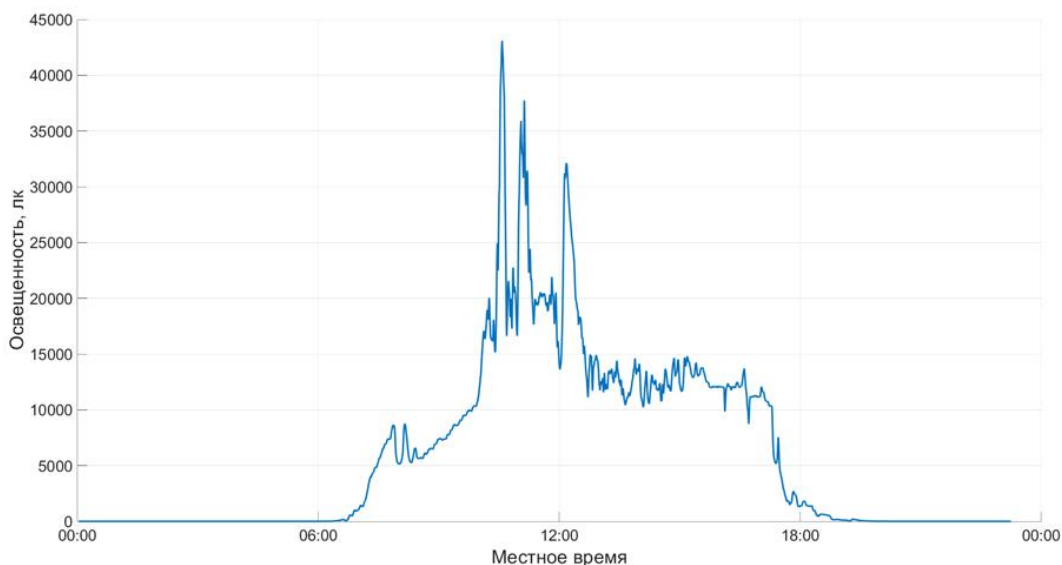


Рис. 3. Зависимость интенсивности солнечного света от времени суток

ВЫВОДЫ

Стоит отметить, что датчик обладает ограниченным диапазоном измерений, составляющим всего до 65535 Lx, в то время как максимальная солнечная освещенность при ясном небе может достигать 100 kLx [4]. Эта проблема может быть решена путем применения полупрозрачного стекла или поляризаторов, что позволит расширить диапазон измерений.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Modular Sun Sensor System for Orbital Attitude Determination [Электронный ресурс]. URL: https://honors.libraries.psu.edu/files/final_submissions/1553 (дата обращения: 30.03.2024).
2. A photodiode based miniature sun sensor / Lü X., Tao Y., Xie K. [et al.] // Measurement Science and Technolog. V.28. N.5. 2017. P. 1–8.
3. Allgeier S. E., Mahin M., Fitz-Coy N. G. Design and Analysis of a Coarse Sun Sensor for Pico-Satellites // AIAA Infotech, Aerospace
4. Ландсберг Г.С. (ред.) Элементарный учебник по физике: Учебное пособие в 3-х т. Т. 3. Колебания и волны. Оптика. Атомная и ядерная физика. 10-е изд., перераб. М.: Наука. Главная редакция физико-математической литературы. 1986. С. 656.

ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА ДЛЯ ДАЛЬНЕЙШЕГО СЕМАНТИЧЕСКОГО АНАЛИЗА

М.Е. Бородина, В.А. Кудинов

Курский государственный университет, г. Курск, Россия

E-mail: borodina_me@kursksu.ru

В данной статье предложена общая для всех методов семантического анализа текста на естественном языке технология предобработки и обработки данных. Обоснована необходимость в разработке индивидуального подхода к реализации этапов технологии на основе задач семантического анализа.

Ключевые слова: предобработка текста; обработка естественного языка.

ВВЕДЕНИЕ

Современное информационное общество и повсеместная цифровизация способствуют росту количества текстовой информации, требующей обработки. Одним из важных направлений анализа текста является семантический анализ, расширяющий возможности систем автоматической обработки текста (АОТ). Как правило, перед исследователем стоит выбор метода семантического анализа, зависящий от решаемых задач. Однако эта зависимость распространяется и на технологию обработки данных. Рассмотрим подробнее этапы работы с данными, предшествующие непосредственному применению метода семантического анализа текста на естественном языке (ЕЯ), чтобы показать эту зависимость.

Понимание естественного языка компьютером является сложной задачей, так как он обладает свойствами, которые усложняют применение вычислительных подходов [1]. Основным элементом письменной речи является дискретная единица – символ. Символы формируют слова, составляющие предложения. Символы, слова и предложения получают интерпретации в нашей голове. При этом переход между интерпретациями невозможно осуществить с помощью «простой» математической операции, как, например, при переходе от красного цвета к зеленому при кодировке. При работе с текстом переход «красный» → «зеленый» потребует справочной информации или словаря. Все это говорит о разном «восприятии» текста компьютерной системой и человеком, поэтому после формирования массива текстовых данных следует их предобработка и обработка (см. рисунок ниже).

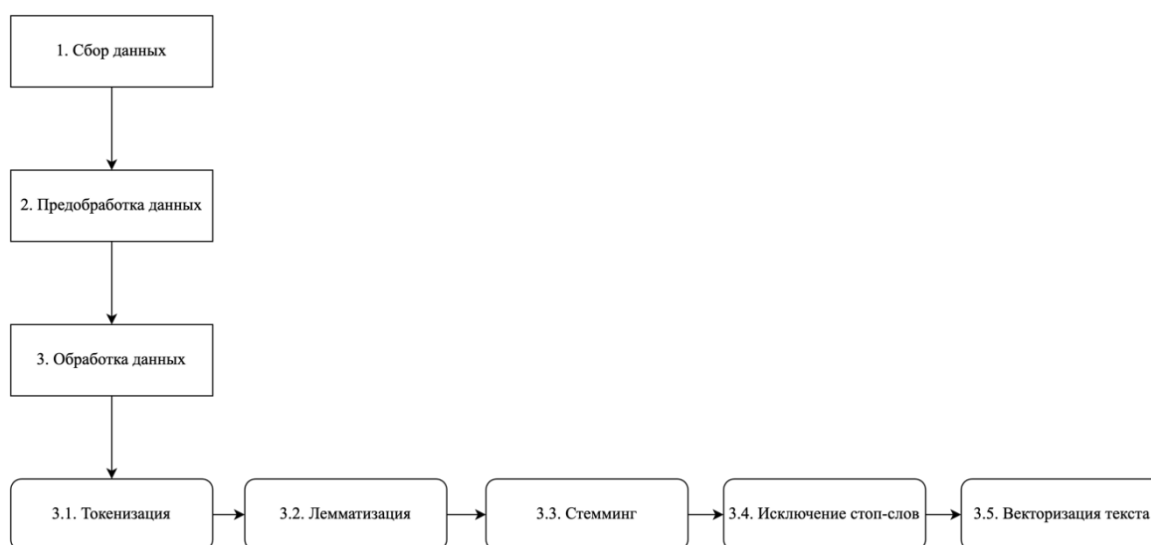
1. СБОР ДАННЫХ

Данные, на результатах обработки которых будет базироваться метод семантического анализа, можно разделить на субъективные и объективные.

К **субъективным данным** относятся результаты эксперимента, позволяющего описать лингвистическое явление, смоделировать человеческое мышление, например, опрос носителей русского языка с целью соотношения к той или иной возрастной группе мужчин различного возраста [2].

К **объективным данным** относят подобранную и обработанную по определённым правилам совокупность текстов, используемых в качестве базы для исследования языка [3].

Многие исследования в области языкознания строятся на основе обработки именно субъективных вероятностных оценок, получаемых исследователем от информанта [2], что отвечает цели семантического анализа текста ЕЯ – приближению машинного «понимания» текста к человеческому восприятию семантики текста.



Технология обработки данных в текстовой форме

2. ПРЕДОБРАБОТКА

В процессе предобработки цифрового текста на естественном языке следует учитывать кодирование, используемое в конкретном документе, поэтому корпус должен быть приведен к единой кодировке [4].

Приведение текста к единому регистру и исключение шума в тексте также необходимы для его подготовки к последующей работе. Наличие гиперссылок, расширений файлов, аббревиатур, повторяющихся знаков

препинания, специальных символов, их комбинаций, чисел требует особого вида предобработки [5].

3. ОБРАБОТКА ДАННЫХ

Обработка естественного языка – это область исследований в компьютерных науках и искусственном интеллекте (ИИ), занимающаяся интерпретацией, манипулированием, пониманием естественных языков. Обработка включает в себя перевод естественного языка в числовые данные, используя которые можно получать необходимую информацию об окружающем мире с использованием возможностей вычислительной техники [6].

Рассмотрим этапы обработки текста на естественном языке более подробно [1].

3.1. ТОКЕНИЗАЦИЯ

Токенизация – это процесс разбиения фразы, предложения, абзаца или всего текстового документа на более мелкие единицы, например, отдельные слова или термины [7].

На этапе токенизации могут возникнуть различные неоднозначности. Например, одно и то же слово может быть написано по-разному (слитно, раздельно или через дефис) [5]. Примером таких слов могут быть слова, заимствованные в русский язык из других языков (офлайн, оффлайн или офф-лайн).

Другой проблемой является то, что слова объединяются в словосочетания и предложения по-разному и, следовательно, определение границ предложений тоже может быть связано с определенными трудностями, так как точки не всегда являются символом окончания строки. Они могут встречаться внутри предложений, например, после сокращенных слов или же использоваться в качестве специального символа в техническом документе [5].

3.2. ЛЕММАТИЗАЦИЯ

Лемматизация – процесс приведения слова к нормальной (словарной) форме [1]. С одной стороны, однокоренные слова воспринимаются в качестве одного слова, что делает текстовую модель более общей, простой, с другой – не учитывается различие смысла однокоренных слов [6].

3.3. СТЕММИНГ

Стемминг является более «грубым» процессом, чем лемматизация, потому что заменяет слово на его основу, которая может не совпадать с

корнем. Проблемой этого процесса является несовершенство работы стеммеров и приведения к единому смыслу омографов.

3.4. ИСКЛЮЧЕНИЕ СТОП-СЛОВ

Стоп-слова – это слова, не несущие в себе содержательной информации о смысле фразы. Например, предлоги, частицы, вводные обороты речи. Их исключение сокращает шум, содержащийся в корпусе.

3.5. ВЕКТОРИЗАЦИЯ ТЕКСТА

В простейшем варианте слова текста заменены номерами позиций в частотном словаре [6]. Наиболее эффективными и распространенными методами являются Bag Of Words и TF-IDF.

Таким образом, предлагаемая технология позволяет не только получить готовый для компьютерной обработки текст, но и повысить качество обрабатываемых данных за счет эффективного исключения шума, так как содержит этапы лемматизации и стемминга.

Для каждого из этапов технологии рассмотрены варианты реализации, выбор которых обеспечивает соответствие обрабатываемых данных задачам семантического анализа.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Goldberg Y.* Neural Network Methods in Natural Language Processing // San Rafael, California (USA): Morgan & Claypool Publishers, 2017. С. 287.
2. *Пиотровский Р.Г., Бектаев К.Б., Пиотровская А.А.* Математическая лингвистика. Учеб. пособие для пед. институтов // М.: Высшая школа, 1977. С. 383.
3. *Лаврентьев А.М.* Корпусная лингвистика: идеология, методы, технологии // Сибирский филологический журнал, 2004. № 3-4. С. 122.
4. *Цитильский А.М., Иванников А.В., Rogov И.С.* NLP – обработка естественных языков // Научно-образовательный журнал для студентов и преподавателей «StudNet». 2020., № 6. С. 467–475.
5. *Боярский К.К.* Введение в компьютерную лингвистику. Учебное пособие. // СПб: НИУ ИТМО, 2013. С. 72.
6. *Хобсон Л., Ханнес Х., Коул Х.* Обработка естественного языка в действии // СПб: Питер, 2020. С. 576.
7. *Вашкевич Е.К.* Токенизация в NLP // Инфокоммуникации: сборник тезисов докладов 56-ой научной конференции аспирантов, магистрантов и студентов БГУИР, Минск, 18 – 20 мая 2020 г. Минск: БГУИР, 2020. С. 67–68.
8. *Прошина М.В.* Современные методы обработки естественного языка: нейронные сети // Экономика строительства. 2022. № 5. С. 27–42.

РАСПОЗНАВАНИЕ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ - «СЧАСТЛИВЫЙ ЧЕЛОВЕК»

Чжуню Го, В.С. Садов

Белорусский государственный университет, Минск, Беларусь

E-mail: guozhongyou1@gmail.com

В данной работе исследованы методы использования нейронных сетей для распознавания эмоции «счастливый человек» по изображению лица. Мы использовали набор данных FER-2013, который содержит 28709 примеров для обучения нейронной сети и 3589 примеров для ее тестирования. Обучающая и тестовая выборки включают семь базовых эмоциональных категорий (злость, отвращение, страх, радость, грусть, удивление, нейтральность). Использована сверточная нейронная сеть (CNN) для извлечения особенностей лица, Точности распознавания эмоций составила 69%.

Ключевые слова: распознавание эмоций; базовые эмоции; глубокое обучение; нейронная сеть; модель.

ВВЕДЕНИЕ

Распознавание эмоций по лицу – это область, использующая технологии нейронных сетей для анализа и распознавания выражений лица. С развитием искусственного интеллекта нейронные сети достигли значительных успехов в области компьютерного зрения, и распознавание эмоций по лицу стало распространенным направлением исследований.

Цель распознавания эмоций по лицу заключается в точном определении эмоционального состояния человека путем анализа физиологических характеристик лица, и определения по ним качественного и количественного состава отображаемых базовых эмоций.

Полученные данные позволяют спрогнозировать поведенческую линию человека, оценить его способность к тем или иным действиям. Эта технология имеет широкое применение во многих областях, включая взаимодействие человека с компьютером, эмоциональные вычисления, маркетинговые исследования и т. д.

ИСПОЛЬЗУЕМАЯ БАЗА ДАННЫХ

В данном исследовании использовалась база данных FER-2013 для анализа изображений, посредством сверточной нейронной сети. Размер изображений в этом наборе данных составляет 48x48 пикселей, и они представлены только черно-белыми изображениями с одним цветовым каналом. Эта база данных содержит 28709 примеров в обучающем наборе и 3589 [1] примеров в тестовом наборе. Оба набора данных включают изображения лиц разных людей, выражающих 7 базовых эмоций (гнев, отвращение, страх, радость, грусть, удивление, нейтральность) [1].

В табл. 1 приведен количественный и качественный состав базы данных FER-2013, а на рис. 1 примеры фотографий с выражениями основных базовых эмоций.

Таблица 1

Наполнение базы (выборки данных) FER-2013

Базовые эмоции	Train	Test
Радость	7215	1774
Нейтральность	4965	1223
Грусть	4830	1247
Страх	4097	1024
Злость	3995	958
Удивление	3171	831
Отвращение	436	111

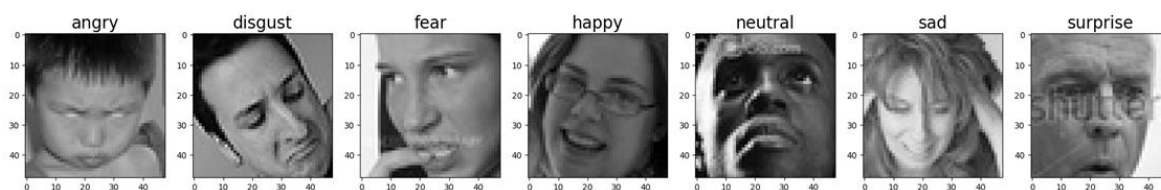


Рис. 1. Примеры образцов из базы данных FER-2013

МЕТОД ИССЛЕДОВАНИЙ

В данном исследовании использовалась нейронная сеть архитектуры ResNet50V2 для анализа данных. Основная идея ResNet50V2 [2] заключается в использовании блоков с остаточными соединениями (Residual Connection) для решения проблемы затухания и всплеска градиента в процессе обучения глубоких нейронных сетей. В традиционных сверточных нейронных сетях с увеличением количества слоев градиент постепенно уменьшается, что делает обучение сложным. Однако благодаря остаточным соединениям сеть позволяет градиенту быстрее распространяться, что эффективно решает проблемы его затухания и всплеска.

Архитектура сети ResNet50V2 включает в себя сверточные слои, слои пакетной нормализации (Batch Normalization), функции активации и блоки с остаточными соединениями. Блок с остаточными соединениями является ключевым компонентом ResNet и состоит из нескольких сверточных слоев и остаточного соединения. Внутри каждого блока обычно находятся два сверточных слоя размером 3x3 и одно остаточное соедине-

ние, а также слои пакетной нормализации и функции активации для улучшения выразительности сети. На рис. 2 и в табл. 2 приведены графики функции потерь и точности при обучении сети на наборе данных из базы FER-2013

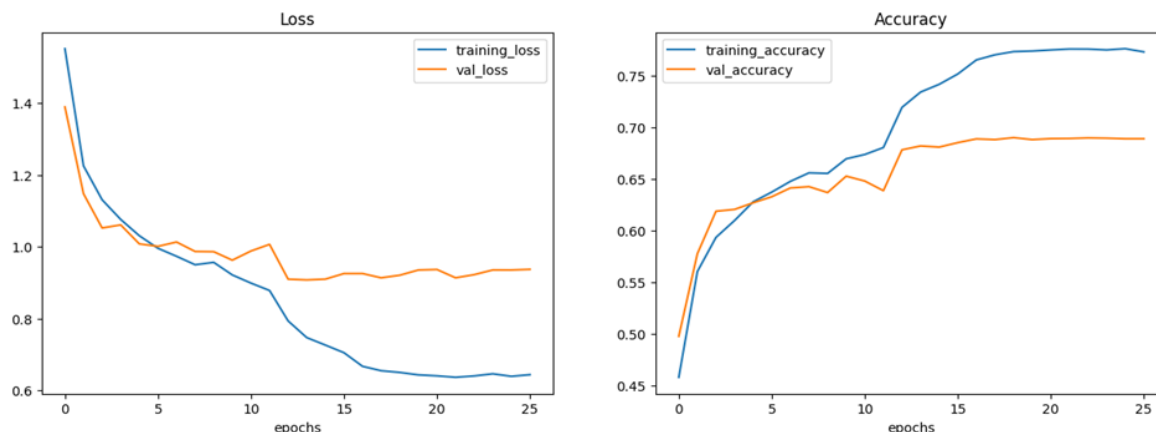


Рис. 2. Графики функции потерь и точности при обучении сети

Test Loss: 0,91942; Test Accuracy: 69,03%

Таблица 2

Результаты оценки базовых эмоций из базы данных FER-2013

Эмоции	Precision	Recall	f1-score	Support
Злость	0,62	0,64	0,63	958
Отвращение	0,78	0,56	0,65	111
Страх	0,56	0,45	0,5	1024
Радость	0,88	0,88	0,88	1774
Нейтральность	0,62	0,71	0,67	1233
Грусть	0,56	0,57	0,57	1247
Удивление	0,8	0,79	0,79	831
accuracy			0,69	7178
macro avg	0,69	0,66	0,67	7178
weighted avg	0,69	0,69	0,69	7178

Для идентификации эмоционального состояния «счастливый человек» по распознанным базовым эмоциям будем использовать следующую модель [3]:

Счастье=($K_{\text{радость}} + K_{\text{удивлен.}} + K_{\text{нейтральн.}}$)/($K_{\text{грусть}} + K_{\text{страх}} + K_{\text{отвращен.}} + K_{\text{гнев}}$) ≥ 3 ,
где K – процентное содержание соответствующей базовой эмоции.

Результаты анализа фотографий, приведенных на рис. 3, на наличие среди них изображения лиц с выражением счастья приведен в табл. 3.

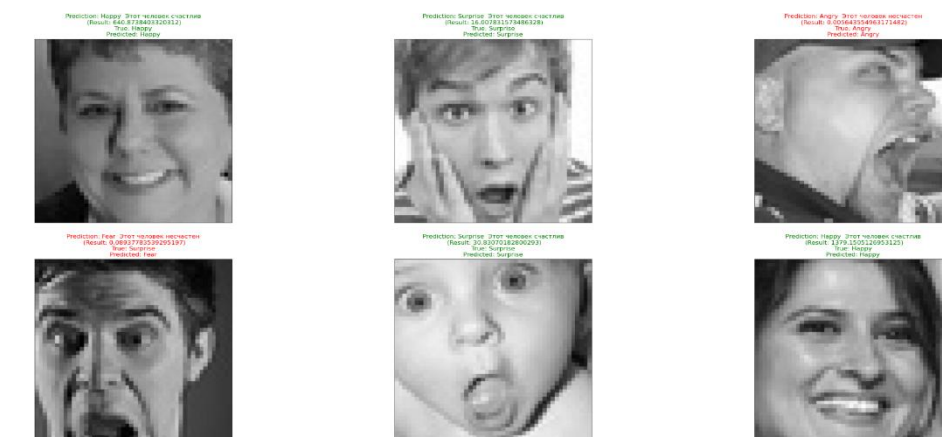


Рис. 3. Тестовая выборка с изображениями лиц счастливого человека (с пояснениями зеленого цвета)

Таблица 3
Результаты анализа тестовой выборка на наличие лиц счастливых людей

№ рис.	Счастливы?	Результат анализа по модели	Оценка эмоции в базе данных	Оценка эмоции разработанной системой
1	да	640,87	Радость	Радость
2	да	16,00	Удивление	Удивление
3	нет	0,005	Злость	Злость
4	нет	0,089	Удивление	Страх
5	да	30,83	Удивление	Удивление
6	да	1379,15	Радость	Радость

ВЫВОДЫ

В нашем исследовании мы использовали ResNet50V2 для распознавания эмоций на лицах, и с помощью оптимизации дизайна сети, обучающей стратегии и использования набора данных FER-2013, был достигнут удовлетворительный результат распознавания – 69,03%.

Эмоциональное состояние «счастливый человек» можно определить по совокупности выявленных базовых эмоций в соответствии с моделью [3].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Kaggle. FER-2013 [Электронный ресурс]. URL: <https://www.kaggle.com/datasets/msambare/fer2013>.
2. He, K., Zhang, X., Ren, S., & Sun, J. Identity Mappings in Deep Residual Networks. In European Conference on Computer Vision, 2016.
3. Фредриксен, Б. Сила позитивных эмоций: Новейшие исследования раскрывают соотношение 3:1, которое изменит вашу жизнь/ Нью-Йорк MJF: Crown, 2009.

РЕАЛИЗАЦИЯ УЧЕБНОГО КОМПЛЕКСА ДЛЯ ИЗМЕРЕНИЯ ХАРАКТЕРИСТИК МАССИВОВ ЛАЗЕРНЫХ ПРОБОВ В СТЕКЛЕ

С.С. Гринько, С.С. Индюков, М.В. Лобанок, О.Р. Людчик

Белорусский государственный университет, Минск, Беларусь

E-mail: Lyudchik@bsu.by

На базе спектрофотометра *S100* и микроконтроллера *Atmel328* собран учебный стенд для измерения оптических характеристик. Проведены исследования влияния плотности записи оптических пробов на пропускание оптического стекла. Наносекундными лазерными импульсами (излучаемыми лазерной системой *Nd:YAG*, $\lambda=1064$ и $W_{\text{имп}}=10-12$ мДж) в оптическом стекле марки К8 сформированы массивы дефектов с разной степенью заполнения и различным количеством слоёв.

Ключевые слова: микроконтроллеры; лазерная обработка; оптические свойства; наносекундное облучение; обработка спектра.

Надёжное долгосрочное хранение и многократное считывание информации является одной из важнейших задач при создании архивов и дата-центров. Концепция долговременной памяти прорабатывается долгое время, однако существенных успехов в записи большого объёма информации удалось достичь только при развитии технологий лазерного взаимодействия с материалами и систем оптического считывания информации. В ряде работ сообщается о возможности записи и хранения информации в кварцевом стекле – технологически простом и бюджетном материале [1,2]. Однако, остаётся ряд не решённых задач, связанных с развитием физико-технологических режимов записи и чтения информации в стекле. Таким образом, создание и стандартизация систем измерения оптических характеристик для прозрачных сред с массивами дефектов является важной задачей для развития технологии долговременной энергонезависимой памяти. С другой стороны, подобная система может быть использована и для проведения рутинных измерений оптических свойств гетероструктур [3]. В работе представлено аппаратно-программное решение измерения оптических характеристик на базе спектрофотометра *S-100*, которое позволяет исследовать зависимости рассеяния проходящего излучения от геометрических параметров дефектов и количества их слоёв.

Массивы дефектов формировали в оптическом стекле марки К8 размером 80x80x10 мм. Лазерную обработку проводили с помощью лазерной системы *Nd:YAG*, $\lambda=1064$ и $W_{\text{имп}}=8-10$ мДж. Интенсивность лазерного излучения контролировалась с помощью *Thorlabs PM100USB*. В работе изучались массивы, приведённые на рис. 1а и содержащие от 1 до 6 слоёв дефектов. Первый слой массива дефектов формировался в 2 мм от ниж-

него края стекла. Дефекты в плоскости слоя находятся в центрах равносторонних треугольников на расстоянии 650 мкм. Они были созданы с разными энергиями накачки 8-8,6 Дж что соответствует энергиям генерации 12,2-19,5 мДж. Расстояние между соседними слоями дефектов задавалось таким образом, чтобы расстояние между центрами дефектов составляло не 1мм.

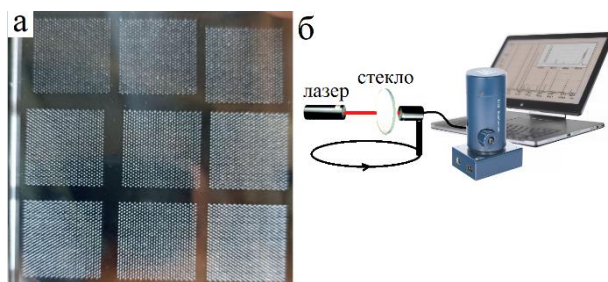


Рис. 1. *а* – Массивы микродефектов, сформированных с разными энергиями накачки (8, 8,3, 8,6 Дж) и количеством слоёв (1-3 слоя); *б* – Схема измерительной установки интенсивности пропускания с угловым разрешением, используемой для измерения рассеяния на дефектах.

Принципиальная схема измерения приведена на рис. 1б. На столе размещены лазер с длиной волны $\lambda \sim 680$ нм, мощностью 0,6 мДж и неподвижный предметный столик с закреплённым образцом, а также подвижный световод спектрофотометра *S100*. Для проведения экспресс-измерений состояний оптических пробоев реализована также схема с неподвижным предметным столиком и независимыми подвижными камерой и лазером с длиной волны $\lambda \sim 532$ нм и линзой для линейной развёртки. Управление скоростью и направлением подвижной частью осуществляется с помощью микроконтроллера Atmel328 и подключённых к нему драйверов шагового двигателя. Массив спектральных данных спектрометра *S100* и данные с камеры поступают на компьютер. Последующая обработка и анализ данных происходит с помощью *MATLAB*.

Обработанные экспериментальные данные зависимости проходящего излучения от количества слоев и угла падения на массив оптических пробоев представлены на рис. 2. Отмечается, что при наличии одного слоя интенсивность света уменьшается на 20%.

При увеличении количества слоёв дефектов интенсивность проходящего излучения снижается в 2 раза. При этом в многослойной системе наблюдается хорошая равномерность (разница не более 5%), поскольку слои формировались строго друг над другом, таким образом, чтобы между областями пробоя оставались промежутки без точек рассеяния. Из зависимости интенсивности (рис. 2а) следует, что варьирование энергией лазерных импульсов приводит к изменению рассеяния излучения.

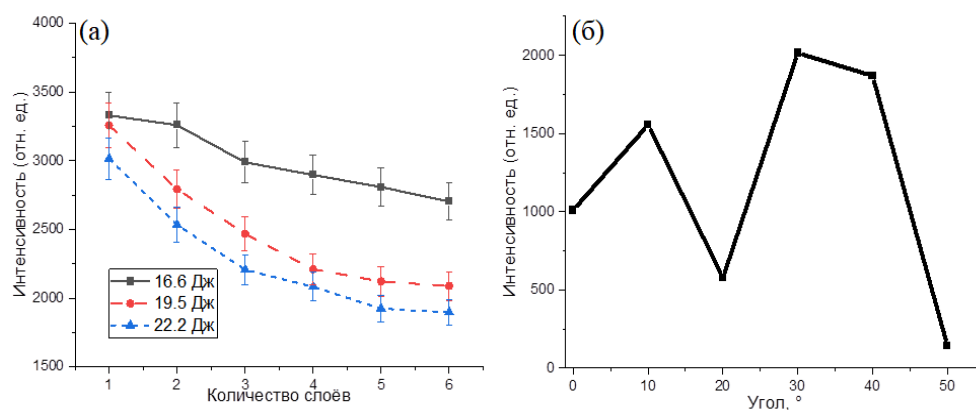


Рис. 2. Зависимость интенсивности проходящего излучения: *а* – от количества слоёв с дефектами; *б* – от угла прохождения в одном дефекте

Увеличение энергии приводит к изменению структурных характеристик массивов дефектов, а именно к увеличению размера дефектов от 0,2 до 0,3 мм по осям x , y , а также от 0,4 до 0,6 мм по оси z . Увеличение плотности записи дефектов приводит к снижению интенсивности пропускания, а также к изменению равномерности рассеяния по площади массива (снижение разницы интенсивности пропускания излучения до 5%). С помощью измерительной системы были проведены измерения рассеяния в зависимости от стороны, с которой излучение попадает на дефект. Для этого в стеклянном цилиндре был сформирован дефект. Результаты измерения коррелируют с известными [4]. В исследовании [4] показано, что геометрия рассеивающей точки оказывает определяющее влияние на оптические характеристики системы. Стоит отметить, что влияние геометрии дефекта на равномерность рассеяния недостаточно изучено.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. 100-Layer Error-Free 5D Optical Data Storage by Ultrafast Laser Nanostructuring in Glass / H. Wang, Y. Lei, L. Wang, M. Sakakura [et al.] // *Laser & Photonics Reviews* 2022, 16, 2100563.
2. C-H, Park, Y. Petit, L. Canioni, S-H. Park. Five-Dimensional Optical Data Storage Based on Ellipse Orientation and Fluorescence Intensity in a Silver-Sensitized Commercial Glass // *Micromachines*. 2020; 11(12):1026.
3. М.В. Лобанок, А.И. Мухаммад, П.И. Гайдук. Структурные и оптические характеристики гетероструктур SiC/Si, полученных методом быстрой вакуумно-термической карбидизации кремния // *Журнал прикладной спектроскопии*. 2022. №2 (89). С. 204-209.
4. T. Alqurashi, A. Alhosani, M. Dauleh, A.K. Yetisen, H. Butt. Laser inscription of pseudorandom structures for microphotonic diffuser applications // *Nanoscale*, 2018; 10: 7095-7107.

ОПТИМИЗАЦИЯ АЛГОРИТМА СТЕГАНОГРАФИЧЕСКОГО СКРЫТИЯ СЛУЖЕБНОЙ ИНФОРМАЦИИ В ПРОСТРАНСТВЕННОМ ПРЕДСТАВЛЕНИИ КАРТОГРАФИЧЕСКИХ ИЗОБРАЖЕНИЙ

Д.М. Заблоская, В.С. Садов

Белорусский государственный университет, Минск, Беларусь
E-mail: dasaz659@gmail.com

В данной работе рассматриваются вопросы стеганографического сокрытия служебной (закрытой) информации в цифровых картах (контейнерах), которые могут находиться в открытом доступе. Скрытие информации осуществляется на основе алгоритма Куттера-Джордана-Боссена дополненного блоком интеллектуального отбора пригодных для стеганографических модификаций пикселей контейнера, осуществляемого в соответствии с пространственным распределением яркостно-цветовой неоднородности его пикселей и объема встраиваемых данных.

Ключевые слова: цифровые картографические изображения; неоднородность изображений; стеганография.

Введение. Картографические изображения подразделяются на общедоступные и конфиденциальные, доступ к которым имеют определенный круг лиц. Стеганография предоставляет надежный и удобный для использования метод хранения и транспортировки таких карт, путем незаметного сокрытия секретных объектов на них в элементах цифрового представления общедоступных карт (контейнеров) [1].

Особенностью стеганографии является установления баланса между стеганографической емкостью контейнера (объемом встраиваемой информации) и незаметностью встраивания [2]. Эта задача требует, в первую очередь, анализа пространственной и яркостно-цветовой неоднородностей изображений-контейнеров. У неоднородных участков изображения выше стеганографическая емкость, в отличие от однородных, и выше незаметность встраивания. То есть для наилучшего встраивания больше подходят неоднородные участки контейнеров [1].

Стеганографический алгоритм, который представлен в данной работе, основан на методе Куттера-Джордана-Боссена. Бит секретного сообщения встраивается в канал синего цвета контейнера путем модификации яркости его пикселей на определенное значение, которое рассчитывается индивидуально для каждого пикселя с учетом объема скрываемых данных, требуемой степени незаметности их встраивания и яркостно-цветовой неоднородности контейнера.

Алгоритм встраивания информации. Алгоритм состоит из шагов:

- Построение карты неоднородности контейнера. Для этого изображение-контейнер сканируется окном (блоком) 5x5 пикселей с шагом 1 пиксель и определяются средние значения яркости пикселей в блоках в соответствии с выражением:

$$avg = (\sum_{i=1}^{25} x_i) / 25, \text{ где } x - i\text{-ый элемент блока.}$$

- Определяется разность по модулю между центральным пикселем блока и средним значением яркости пикселей в этом блоке:

$$diff = |c - avg|, \text{ где } c - \text{центральный элемент, } avg - \text{средняя яркость блока.}$$

- Если эта разность, которую назовем силой неоднородности, находится в заданном диапазоне значений, то координата данного пикселя и значение его яркости заносится в карту неоднородности контейнера.

- Оценивается полученная величина стеганографической емкости контейнера. Если она меньше объема скрываемых данных, то корректируется величина силы неоднородности в сторону уменьшения.

- кодирование бит сообщения:
 - если встраиваемый бит скрываемых данных - «0», то из значения яркости центрального пикселя вычитается разность $diff$. В случае, если центральный пиксель примет, при этом, отрицательное значение - приравниваем его к нулю.
 - если встраиваемый бит скрываемых данных - «1», то к значению яркости центрального пикселя прибавляется разность $diff$. При этом, максимальное значение центрального пикселя, ограничивается величиной - 255.

Результаты встраивания данных. В данном исследовании в качестве контейнера, используется изображение карты местности, размером 2369×2562 пикселей, представленное на рисунке 1.

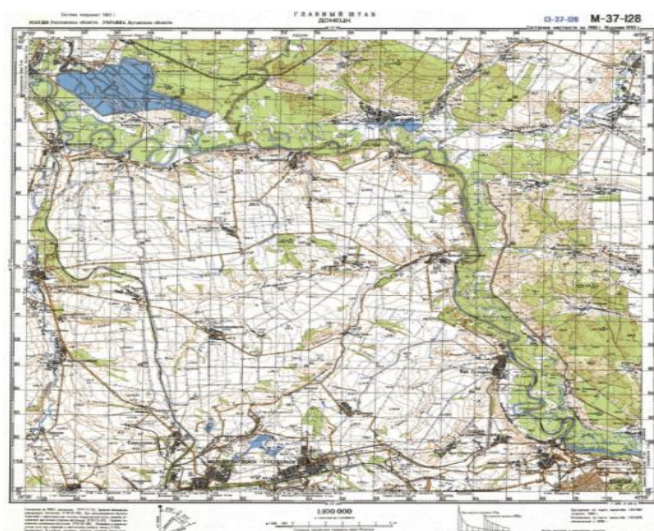


Рис. 1. Картографическое изображение.

Результаты оценки визуальных искажений контейнера по величине PSNR при встраивании графических данных приведены в табл. 1.

Таблица 1

Оценка по величине PSNR заполненных контейнеров

Диапазон неоднородностей используемых блоков	Пиковое отношение сигнал-шум при встраивании графических данных, дБ
0 – 10	57,8258
0 – 20	51,4697
0 – 30	47,5545
0 – 40	44,9591
0 – 50	42,8759
0 – 60	41,4128
0 – 70	39,6954
0 – 80	39,2893

Из табл. 1 видно, что приемлемое значение PSNR обеспечивается при силе неоднородности 10 – 60. Минимальное значение PSNR принимается равным 40 дБ.

Для извлечения информации из заполненного контейнера необходимо знание заданного диапазона силы неоднородности и длины встроенных данных в битах. По этим данным строится карта неоднородности полученного заполненного контейнера. Далее аналогично алгоритму встраивания выделяются на этой карте блоки 5х5 пикселей и для каждого блока по соотношению величины яркости его центрального пикселя и средней яркости всех пикселей в блоке делается вывод о значении информационного бита, встроенного в анализируемом блоке.

В силу того, что карта неоднородности заполненного контейнера строится по измененным в процессе встраивания пикселям, при извлечении встроенной информации возможно ложное распознавание значений встроенных бит (табл. 2).

Таблица 2

Ошибки извлечения текстовой информации

Диапазон неоднородностей используемых блоков	Процент ошибок извлечения текстовой информации
10 – 20	73,08%
20 – 30	10,09%
30 – 40	0%
40 – 50	0%
50 – 60	0%

Из табл. 2 видно, что наилучший, по минимуму ошибки, диапазон силы неоднородности блоков, используемых контейнеров, составляет 30 – 60.

Этот диапазон является оптимальным и для случая встраивания графической информации. Однако, встраиваемое изображение, при этом, не превысит размер – 177х177 пикселей. Более того, этот диапазон, можно расширить до значений 23 – 67, при этом ошибка извлечения данных не превысит 0,38%, а допустимый размер встраиваемых изображений составит 206х206 пикселей. При таком диапазоне неоднородностей блоков контейнера извлеченное сообщение будет выглядеть как показано на рис. 2.

Оригинальное изображение Восстановленное изображение



Рис. 2. Сравнение оригинала сообщения и извлеченного сообщения

Вывод. В стеганографии необходимо следить за балансом между незаметностью встраивания и стеганографической емкостью контейнера.

Для встраиваемых данных текстового вида лучше использовать блоки контейнера с силой неоднородности 30 – 60. При таком диапазоне удастся встроить примерно 19 страниц русского текста, при этом ошибок извлечения совсем не будет.

При встраивании изображений возможно использовать блоки контейнера с силой неоднородности 23 – 67. В таком случае становится возможным встроить изображение размером до 206×206 пикселей. Стеганографическую емкость контейнера можно увеличить практически в двое, либо втрое задействовав остальные цветовые каналы картографического контейнера.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Садов. В.С. Компьютерная стеганография: учеб. пособие / В.С. Садов. Минск: РИВШ, 2014. С. 172 с.: ил.
2. Конахович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. Киев: МК Пресс. 2006.

СИСТЕМА ДИСТАНЦИОННОГО МОНИТОРИНГА БАРОМЕТРИЧЕСКИХ И МЕТЕОРОЛОГИЧЕСКИХ ПАРАМЕТРОВ ОКРУЖАЮЩЕЙ СРЕДЫ

А.А. Камоцкий, С.С. Гринько, А.С. Семенович, М.В. Лобанок

*Белорусский государственный университет, Минск, Беларусь E-mail:
aliaksandr.kamotski@gmail.com*

На базе микроконтроллера ESP32 предложено аппаратно-программное решение для дистанционного измерения барометрических и метеорологических параметров атмосферы. В тестовом режиме получены и обработаны данные с узлов в комнатных и уличных условиях.

Ключевые слова: микроконтроллеры; метеостанция; Wi-Fi; Bluetooth.

Систематическое измерение параметров окружающей среды и их анализ является важной экологической задачей. Существующие доступные метеостанции не обладают возможностями масштабирования, сбора и анализа данных. Помимо этого, существующие системы с беспроводными датчиками для уличного применения обладают рядом недостатков, связанных с ограничением количества подключаемых датчиков и их быстродействием, и достоверностью значений в реальном времени [1].

Аппаратно-программная платформа для уличного и комнатного использования должна обеспечивать возможность подключения разнообразных первичных датчиков, измеряющих барометрические и метеорологические параметры окружающей среды с использованием бесконтактных методов измерения [2]. В данной работе была разработана масштабируемая платформа, в состав которой входят измерительные узлы на базе микроконтроллера ESP32 (рис. 1). При этом реализуется возможность масштабирования измерений: система может состоять как из одного узла, так и из множества, оборудованных различными датчиками, в том числе набором газовых датчиков и дозиметра для детектирования загрязнения окружающей среды.

Для организации корректной работы множества узлов на одном объекте необходимо подключение локального сервера вблизи объекта измерения (комната, здание, машина). Облачное хранилище в данной платформе выступает в роле связующего моста между WEB-сервером и базой данных, в которой хранятся данные объекта измерения. Еще одна задача этого моста – это временное хранилище информации с различных локальных серверов, расположенных в разных объектах измерения.

Все измерительные узлы связаны лишь беспроводными каналами связи, поэтому очень важна взаимная синхронизация «внутренних часов»

работе [3] представлены результаты эксперимента по синхронизации импульсно-связанных осцилляторов с использованием оптического канала связи в группе узловых устройств (роботов) расположенных в матрице 16x4. Каждый узел содержал микропроцессор и поддерживал локальную коммуникацию только с четырьмя ближайшими «соседями», общая синхронизация световых вспышек от всех 64 устройств достигалась лишь за несколько десятков секунд из-за недостатков используемого оптического канала связи синхронизации. В измерительной системе метеостанции подобное медленное решение для синхронизации часов нежелательно, а реализация синхронизации часов в стандарте PTP (IEEE 1588) потребует аппаратной избыточности для каждого конечного узла, что усложняет реализацию измерений в уличных условиях. С другой стороны, постоянное сети устройств на протоколах Wi-Fi и BLE позволяет достичь компромиссного соотношения энергопотребления и эффективностью обмена данными. Огромный объем трафика, генерируемый огромным количеством датчиков и устройств, подключенных к Интернету, должен управляться с помощью совместимых протоколов, особенно тех, которые демонстрируют минимальные характеристики потребления, что является важным критерием для удаленного мониторинга окружающей среды в труднодоступных местах.

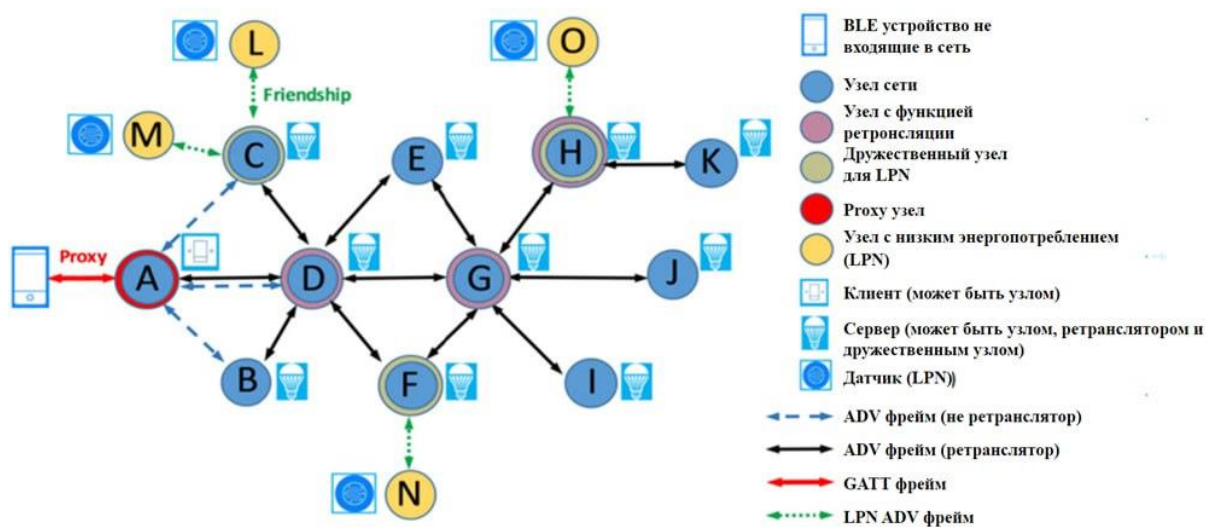


Рис. 1. Структурная схема аппаратной платформы метеостанции

На рис. 2 представлена разработанная структурная схема автономного измерительного узла. В состав электронного устройства вошли электронные подсистемы, успешно опробованные в предыдущих научно-технических проектах с участием автора [3].

Разработанную структурную схему можно разделить на пять основных блоков:

1. Управляющий модуль, главными задачами которого являются обеспечение беспроводной связи устройства с локальным брокером и опрос периферийных датчиков;
2. Модуль индикации (ЖК-дисплей) на измерительном узле, работающем в комнате;
3. Модуль периферийных датчиков, отвечающий за фиксацию показателей окружающей среды;
4. Модуль заряда и контроля параметров аккумуляторной батареи;
5. Модуль TTL преобразователя (UART-COM порт мост).

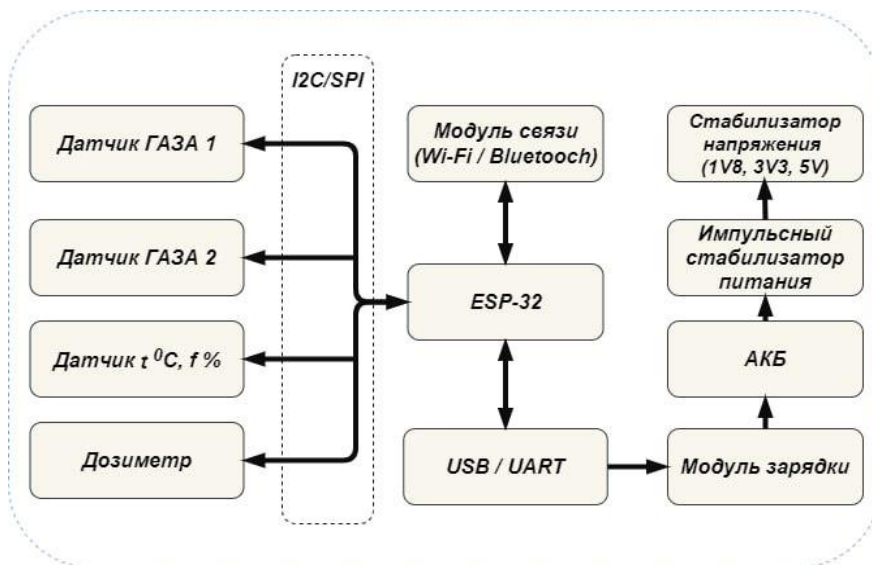


Рис. 2. Структурная схема автономного измерительного узла

Для проведения функциональных тестов созданного узла сети и самой сети разработано тестовое программное обеспечение, которое состоит из двух частей: микропрограммы для узлового устройства и программного модуля для мобильного смарт-устройства управления с возможностями конфигурирования параметров сети и параметров конечного узла (мобильный телефон под управлением ОС Android с пользовательским приложением).

Обработанные экспериментальные данные температуры, влажности и давления, полученные с двух узлов, размещённых в комнатных и уличных условиях представлены на рис. 3. Отмечается, что при измерениях в комнатных условиях на полученные результаты оказало влияние близкое расположение датчика к модулю питания. Использование нескольких датчиков газа серии MQ по соотношению показаний позволяет определить соотношение газов и кажущуюся их концентрацию [4]. Данные с выведенных на уличный измеритель газовых датчиков не приведены.

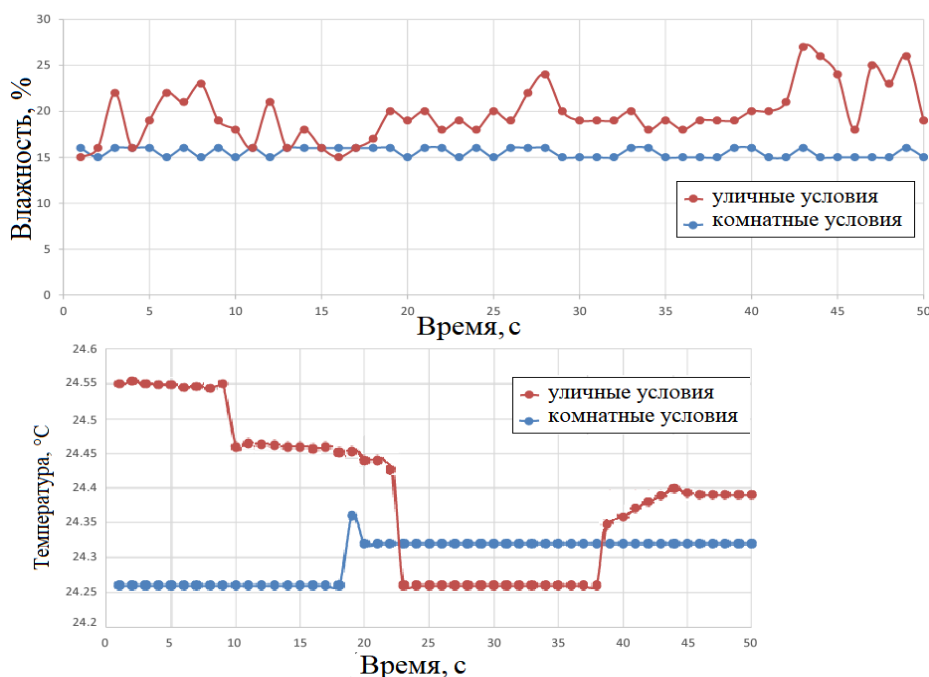


Рис. 3. Экспериментальные измерения температуры и влажности

Из анализа данных следует, что усреднение показания датчиков DHT11 и BME280 позволяют достичь оптимального соотношения инерционности и точности измерений. Для этого, в программной части добавляется задержка в несколько тактов на одном из датчиков. Помимо этого, датчик BME280 передает на сервер показания давления и высоты над уровнем моря, которые могут использоваться для точных барометрических и метеорологических параметров на локальном сервере.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. И.В. Якунина, Н.С. Попов. Методы и приборы контроля окружающей среды. Экологический мониторинг: учебное пособие // Тамбов: Изд-во Тамб. гос. техн. ун-та, 2009. С. 188.
2. А.Г. Дивин, С.В. Пономарев, Г.В. Мозгова. Методы и средства измерений, испытаний и контроля, часть 2 // Издательство ФГБОУ ВПО «ТГТУ», 2012. С. 109.
3. Семенович А.С. Масштабируемый IoT тренажер спортсмена // Сборник научных работ студентов Республики Беларусь «НИРС 2020». С. 143-144.
4. Плевако Е.И., Петров, П.В., Кольчевский, Н.Н. Измерительная система «Colibri» для анализа воздуха на основе беспилотного летательного устройства – квадрокоптера. Квантовая электроника: материалы XIII Международной научно-технической конференции, Минск, 22–26 ноября, 2021. / БГУ, НИИ прикладных физических проблем им. А. Н. Севченко БГУ, Ин-т физики им. Б. И. Степанова НАН Беларуси, Белорусский республиканский фонд фундаментальных исследований; редкол.: М. М. Кугейко (отв. ред.), А. А. Афоненко, А. В. Баркова. Минск : БГУ, 2021. С. 340-342.

ПРИМЕНЕНИЕ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ ДЛЯ ОПТИМИЗАЦИИ КОМПОНЕНТОВ РЕДУКТОРА ПРИ ИЗГОТОВЛЕНИИ С ПОМОЩЬЮ ТРЕХМЕРНОЙ ПЕЧАТИ

Т.Ю. Ким, Е.И. Печковский, А.В. Печковская

Объединённый институт проблем информатики Национальной академии наук

Беларуси, Минск, Республика Беларусь

E-mail: tatyana_kim92@mail.ru, peckovskijevgenij7@gmail.com,

anastasiya.pechkovskaya@gmail.com

В настоящее время трехмерная печать стала важным инструментом прототипирования, ускоряющим внедрение новых технических решений. Данная работа посвящена оптимизации деталей из пластика, изготавливаемых с помощью трехмерной печати. Задача заключается в подборе параметров деталей, позволяющих минимизировать вес комплектующих и расхода пластика в процессе печати. Для этой цели используются генетические алгоритмы и компьютерное моделирование. В работе разработаны функция пригодности и набор критериев для решения задачи оптимизации и обеспечения требуемых характеристик деталей. Метод, предложенный в статье, иллюстрируется на примере оптимизации массы зубчатых колес для цилиндрической зубчатой передачи.

Ключевые слова: Генетические алгоритмы (ГА); компоненты редуктора; оптимизация; масса; аддитивные технологии; эволюционная оптимизация.

Работа выполнена в рамках задания 1.3.1. ГПНИ «Цифровые и космические технологии, безопасность человека, общества и государства», подпрограмма «Цифровые технологии и космическая информатика» и частично при поддержке гранта ГКНТ–Китай Ф22КИТГ-002 «Медицинские роботы на основе человеко-машинного взаимодействия».

ВВЕДЕНИЕ

Целью работы является создание новых подходов к проектированию деталей для изготовления методом трехмерной печати. Работа построена на использовании эволюционных алгоритмов для оптимизации массы детали, а следовательно, и расхода пластика для трехмерной печати, при сохранении прочностных характеристик.

На сегодняшний день аддитивные технологии производства все более востребованы при прототипировании и макетировании, а также при мелкосерийном производстве. Для изготовления компонентов редуктора будет применяться PLA-пластик (англ – Polylactic Acid, полилактид) [1]. Низкая температура плавления такого пластика позволяет обеспечить более высокую точность распечатанной детали.

Изготовление изделий методом трёхмерной печати, равно как и при классических методах, требует оптимизации массы компонентов редук-

тора. Также можно отметить, что многие исследования в области оптимизации конструкции сосредоточены на минимизации массы зубчатой передачи редуктора [2]. Предлагаемый в настоящей работе оптимизационный метод относится к классу эволюционных алгоритмов и базируется на эволюции некоторой популяции особей. Среди эволюционных методов наибольшее распространение получил генетический алгоритм (ГА), предназначенный для оптимизации функций дискретных переменных. К примеру, в [3] представлен процесс автоматизированного проектирования зубчатых передач путём минимизации объёма передач с использованием ГА. Учитывая, что поставленная задача по сути является многокритериальной оптимизационной задачей, целесообразно изучить возможности применения именно ГА для её решения [4].

РЕАЛИЗАЦИЯ ГЕНЕТИЧЕСКОГО АЛГОРИТМА

Для того, чтобы определить работоспособность редуктора при минимальном весе, необходим минимальный перечень характеристик, описывающих требуемые параметры редуктора, куда входит крутящий момент на шестерне и колесе, плотность PLA-пластика (ρ), длина вала шестерни и колеса. Для решения проблемы с помощью ГА была разработана функция пригодности и вычислены оптимальные критерии работоспособности. Была сформирована начальная хромосома, содержащая параметры, которые необходимо оптимизировать. Для этого создаётся вектор X , представленный в следующей формуле:

$$X = [x_1 \ x_2 \ x_3 \ x_4 \ x_5] = [b \ d_1 \ d_2 \ z_1 \ m], \quad (1)$$

где b – ширина рабочего венца, d_1 – диаметр вала шестерни, d_2 – диаметр вала колеса, z_1 – количество зубьев шестерни, m – модуль зацепления.

Далее рассмотрим формирование функции пригодности на примере одноклассового цилиндрического редуктора, который широко используется в общемашиностроительном применении.

Для того, чтобы найти приближенно-точную массу компонентов редуктора, определим основные геометрические размеры цилиндрической прямозубой передачи. Далее вычислим площадь поверхности каждого компонента, затем общую массу компонентов редуктора. Функция пригодности ($f_1(x)$) определяется как суммарная масса исполнительных компонентов редуктора, которую следует минимизировать:

$$[f(x)] = [M_{\text{gearbox}}] = \rho \cdot (V_{\text{gear}} + V_{\text{pinion}} + V_{\text{pinion\_shaft}} + V_{\text{gear\_shaft}}), \quad (2)$$

где V_{gear} и V_{pinion} – объём шестерни и колеса, $V_{\text{pinion\_shaft}}$ и $V_{\text{gear\_shaft}}$ – объём вала для шестерни и колеса соответственно.

Учитывая, что геометрия шестерни достаточно специфична, следующим этапом будет формирование ограничений. Поэтому существуют требования к качеству зубчатых передач. Выдвигаются ограничивающие условия на контактную прочность, на выносливость при изгибе, на статическую прочность вала, на усталостную прочность, которые должны быть удовлетворены и представлены в виде математических функций для лучшего выбора оптимальных результатов. Основные ограничения для компонентов редуктора, рассматриваемого в качестве примера, представлены ниже.

При формировании нового поколения каждый раз производится расчет на предварительные геометрические размеры передачи, определяемые из условий обеспечения работоспособности по основным критериям. Во время работы ГА данные критерии проверяют на соответствие условиям пригодности параметров шестерни, колеса и вала. В случае удовлетворительного результата предварительные размеры принимаются в качестве окончательных.

Сформулируем критерий оптимальности (3-6):

- расчет на контактную прочность

$$g_1(x) = b_1 - Z_H \cdot \sqrt{\frac{2 \cdot T_1 \cdot (u+1) \cdot K_H}{b \cdot D_1^2}} \leq 0, \quad (3)$$

- расчёт на выносливость при изгибе для шестерни

$$g_2(x) = b_2 - \frac{\sigma_{F2} \cdot Y_{F1}}{Y_{F2}} \leq 0, \quad (4)$$

- расчет на статическую прочность при кратковременных перегрузках

$$g_3(x) = b_4 - \frac{K_P \cdot M_C}{\frac{\pi \cdot d_1^3}{32} \cdot \frac{b \cdot h \cdot (2 \cdot d_1 - h)^2}{16 \cdot d_1}} \leq 0, \quad (5)$$

- расчет на усталостную прочность для быстроходного и тихоходного вала

$$g_4(x) = b_5 < \frac{S_\sigma \cdot S_\tau}{\sqrt{S_\sigma^2 + S_\tau^2}} < b_6 \quad (6)$$

ОБСУЖДЕНИЕ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

Полученную по результатам оптимизации массу шестерён для изготовления методом трехмерной печати целесообразно сравнить с массой аналогичных шестерён, изготавливаемых литьём.

По завершении работы алгоритма, все хромосомы в популяции были сведены к лучшей хромосоме. Было найдено 5 основных параметров (см. таблицу), которые удовлетворяют критериям пригодности. Так же можно найти параметры для редуктора по минимуму для экспериментальной трехмерной печати. Средний вес компонентов редуктора составляет 455

грамм. С учетом того, что заполняемость пластиком при трехмерной печати составляет 30%, то масса компонентов редуктора составляет 45% от таковой для литых деталей.

Значения параметров для компонентов редуктора, которые прошли отбор и удовлетворяют все критериям пригодности

b	d ₁	d ₂	z ₁	m	Вес	Вес 30%
13,625	10,277	10,201	20	3,999		
15,571	18,220	20,190	30	3,970		
10	10	10	24	3,98		
31	10,2	12,6	25	3,5		
10	12,270	10	25	3,97		
Средний вес компонентов					455	354

Применённый метод настройки ГА продемонстрировал эффективность в выборе лучших параметров для трехмерной печати. Предложенная функция пригодности, описываемая формулами (2) и критерии условности с помощью формул (3-6) для проверки работоспособности всего механизма позволили найти оптимальные значения для компонентов редуктора.

ВЫВОДЫ

В ходе исследования была смоделирована и подвергнута статическому структурному анализу прямозубая пара зубчатых колес, конструкция которых была оптимизирована для снижения массы при изготовлении методом трехмерной печати из пластика PLA.

Цель исследования была достигнута с помощью ГА при некоторых ограничениях, таких как прочность на изгиб, контактное напряжение, которые должны выполняться и соблюдении всех размерных условий зубчатых колес.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. S. Singh, Ch. Prakash, S. Ramakrishna. Additive Manufacturing. 2020. P. 176.
2. V. Savsani, R.V. Rao, D.P. Vakharia. Optimal weight design of a gear train using particle swarm optimization and simulated annealing algorithms, mech mach theory 45, 2010. P. 531-541.
3. C. Gologlu, M. Zeyveli. A genetic approach to automate preliminary design of gear drives, Computers & Industrial Engineering 57. 2009. P. 1043-1051.
4. Ким Т.Ю., Прокопович Г.А. Оптимизация коэффициентов ПИД-регулятора системы управления движением мобильного робота по цветоконтрастной линии на основе генетического алгоритма // Информатика. Т. 18, № 4. 2021. С. 53-68. DOI: 10.37661/1816-0301-2021-18-4-53-68 Т. 19, № 3.

МОДЕЛИРОВАНИЕ УГРОЗ УСТРОЙСТВАМ ИНТЕРНЕТА ВЕЩЕЙ

А.О. Кириленко, Е.И. Козлова

*Белорусский государственный университет, Минск, Беларусь
E-mail: annakirilenko23@mail.ru*

В работе представлен краткий обзор угроз безопасности устройств интернета вещей. Кратко описан метод моделирования угроз на основе классификации STRIDE, разработанной Microsoft. Проведен анализ угроз операционной системе и программным компонентам «умных часов» на основе модели классификации угроз STRIDE.

Ключевые слова: интернет вещей; безопасность программного обеспечения; моделирование угроз.

ВВЕДЕНИЕ

Интернет вещей (англ. Internet of Things, сокращенно IoT) – это концепция, представленная физическими объектами («вещами»), которые оборудованы датчиками, программным обеспечением и другими технологиями для сбора и обмена данными через Интернет с помощью стандартных протоколов связи. Данная концепция получила широкое распространение и представлена в различных сферах жизни человека: в промышленности, медицине, строительстве, торговле, транспорте и т.д.

Наряду со всеми преимуществами и возможностями, концепция Интернета вещей имеет свои недостатки, одним из которых является обеспечение безопасности данной технологии. Многие устройства IoT разработаны с ограниченными мерами безопасности, что делает их уязвимыми для хакерских атак. Это связано с достаточно большим количеством неразрешенных проблем: отсутствие единой концепции относительно норм и правил в области безопасности, недостаток опытных специалистов по кибербезопасности в сфере Интернета вещей, желание производителей сократить затраты на производство своего продукта, ограничивая некоторые функции безопасности, что влечет за собой негативные последствия с точки зрения обеспечения защиты устройств.

ОБЗОР УГРОЗ ИНТЕРНЕТА ВЕЩЕЙ

Угрозы технологии Интернета вещей можно разделить на три основные группы [1]:

- угрозы, характерные для любой интернет-системы;
- угрозы, характерные только для IoT-устройств;
- угрозы, связанные с аппаратной частью устройств, например, неправильное использование оборудования и компонентов.

Традиционные методы обеспечения безопасности, такие как блокирование открытых портов на устройствах, относятся к первой группе угроз. Например, холодильник, подключенный к Интернету для отправки информации, может использовать незащищенный SMTP-сервер, вследствие чего он может быть подвержен атакам ботнетов. Ко второй группе относятся угрозы, связанные с оборудованием Интернета вещей. Например, некоторые IoT-устройства имеют ограничения, которых нет у обычного компьютерного оборудования, например очень малые размеры, небольшую память, ограничения мощности, что не позволяет применять подходы классической ИТ-защиты, рассчитанные на более крупные и мощные устройства [2].

МОДЕЛИРОВАНИЕ УГРОЗ МЕТОДОМ STRIDE

Моделирование угроз представляет собой процесс выявления и оценки потенциальных угроз для устройств. Это является очень важным компонентом в обеспечении безопасности системы, поскольку моделирование угроз помогает сформировать актуальный вектор действий злоумышленника и разработать стратегии защиты и предотвращения атак.

Одним из способов моделирования угроз является использование модели классификации угроз STRIDE, разработанной Microsoft [3]. Данная модель позволяет классифицировать угрозы, используя 6 категорий:

- *Spoofing (подмена)* – использование злоумышленником методов маскирования для получения доступа к системе.
- *Tampering (вмешательство)* – нарушение злоумышленником целостности данных (изменение данных).
- *Repudiation (сокрытие)* – сокрытие определенных действий, выполненных в отношении системы.
- *Information disclosure (утечка информации)* – нарушение конфиденциальности данных в системе.
- *Denial of Service (D.o.S, атака «отказ в обслуживании»)* – злоумышленник делает систему недоступной или непригодной для использования легальными пользователями.
- *Elevation of privilege (повышение привилегий)* – непривилегированный пользователь может самостоятельно расширить возможности доступа.

ИСПОЛЬЗОВАНИЕ МЕТОДА STRIDE НА ПРИМЕРЕ УМНЫХ ЧАСОВ

Умные часы, на данный момент, являются одним из популярных устройств, используемых людьми, поскольку предоставляют широкий спектр функций благодаря встроенным датчикам, делая часы настоящим

IoT-устройством. Как и любые другие устройства Интернета вещей, умные часы подвержены угрозам безопасности, поэтому используем метод STRIDE для выявления потенциальных угроз.

Архитектура умных часов для исследования угроз безопасности включает в себя такие блоки как Пользователь → Умные часы (Операционная система, Программное обеспечение, Аппаратная часть) → Сервер (Мобильный телефон) → Внешний сервер (мобильный оператор, внешние сервисы, другие сервера и хранилища данных).

Рассмотрим основные угрозы операционной системе и программному обеспечению (прошивке компонентов). Эти компоненты умных часов позволяют управлять устройством, настраивать функции, обрабатывать данные, получаемые от различных физических модулей часов, а также обеспечивают совместимость с мобильными устройствами.

Подмена: 1) слабый пин-код, что может дать доступ к устройству злоумышленнику, 2) программное обеспечение может не соответствовать последней доступной версии от поставщика, вследствие, например недостаточной степени защищенности операционной системы, установленной на часы производителем, либо внешнего вмешательства на этапе запроса обновлений.

Вмешательство: установка вредоносного ПО, которое позволяет обходить различные функции безопасности, получать и изменять данные о пользователе, блокировать некоторые функции устройства.

Соккрытие: злоумышленник, получив доступ к программному обеспечению, может изменять или засорять данные, связанные с последними изменениями в системе, таким образом, скрывая факт совершения действий в отношении устройства.

Утечка информации: собранные на умных часах данные могут отправляться на сервер производителя, например, для обновления ПО или же на другие сервисы для анализа и обработки, но это представляет собой угрозу, в случае если данные попадут в руки сторонним лицам.

Отказ в обслуживании: злоумышленник, может воздействовать на устройство, например, выключать или перезапускать его систему, таким образом вызывая сбой в работе устройства.

Повышение привилегий: если часы используют устаревшую версию ПО, то возможны уязвимости их функционала, чем злоумышленник успешно пользуется и получает привилегированный доступ к системе.

После выявления и классификации угроз, необходимо их минимизировать, что является важным этапом в осуществлении безопасности устройств.

Угроза, связанная со слабым пин-кодом, решается путем применения более надежного и сложного пароля, который исключает использование

личных данных, например, даты рождения или стандартных последовательностей цифр.

Для минимизации угрозы установки вредоносного ПО и утечки данных используются защищенные каналы связи, которые поддерживают стойкое шифрование между устройством и сервером. В таком случае сервер идентифицирует себя посредством предоставления цифрового сертификата, после чего устанавливается зашифрованный канал для обмена данными.

Угроза получения злоумышленником доступа к программному обеспечению возникает из-за того, что некоторые поставщики делают свои микропрограммы общедоступными, вследствие чего злоумышленник без особых проблем может ими завладеть, либо поставщик сам может предоставить прошивку лицу, которое не имеет на это соответствующих прав. Данная проблема решается путем ограничения доступа к микропрограммам для сторонних лиц.

ЗАКЛЮЧЕНИЕ

Стремительное распространение устройств Интернета вещей требует от производителей гарантии безопасности их продукта на всех этапах разработки. Проведенный обзор угроз IoT-устройств показывает, что существует значительное количество неразрешенных проблем в сфере безопасности. Для их решения применяются различные методы, например, моделирование угроз. В данной работе был рассмотрен один из способов моделирования угроз – метод STRIDE. Применение данного метода для анализа безопасности умных часов показывает, что он помогает выделить наиболее критические угрозы, достаточно прост в применении, что делает его доступным для широкого круга специалистов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Е. А. Верещагина, И. О. Капещкий, А. С. Ярмонов.* Проблемы безопасности Интернета вещей. Учебное пособие // М.: Мир науки, 2021. С. 105.
2. Отдельные аспекты угроз безопасности интернета вещей [Электронный ресурс]. URL: <https://na-journal.ru/11-2023-informacionnye-tehnologii/7001-otdelnye-aspekty-ugroz-bezopasnosti-interneta-veshchei-sovremennoe-sostoyanieb> (дата обращения: 17.03.2024).
3. *Ф. Чанцис, И. Стаис, П. Кальдерон, Е. Деирманцоглу, Б. Вудс.* Практический хакинг интернета вещей // М.: ДМК Пресс, 2022. С. 480.

АЛГОРИТМ АНАЛИЗА МИКРОВЫРАЖЕНИЙ ЛИЦА ПРИ ВЫЯВЛЕНИИ ЛЖИ

К.А. Котова, В.С. Садов

*Белорусский государственный университет, Минск, Беларусь
E-mail: ksyu.kotova.79@gmail.com*

В работе предлагается алгоритм анализа микровыражений лица по видеоряду с целью определения по ним истинного отношения коммуницируемых людей к предмету обсуждения. Используется алгоритм извлечения ключевых точек на лице и их классификация с использованием метода случайного леса. Для обучения алгоритма необходимы базы данных по участникам коммуникационного процесса. Данный подход позволяет оценить эмоциональный фон собеседников с точностью 93%.

Ключевые слова: микровыражения на лице; ключевые точки на лице; метод случайного леса; эмоциональный фон.

ВВЕДЕНИЕ

В данной работе решается задача создания алгоритма определения эмоций при анализе микровыражений лица человека, во время произнесения лживой или правдивой информации, и исследования его эффективности. Согласно Полу Экману базовыми эмоциями человека, которые существенно изменяют выражение лица, являются: удивление, страх, отвращение, гнев, радость, печаль и нейтральное состояние [1].

Эмоции и проявление лжи – это сложные психофизиологические категории, и их взаимосвязь может быть нелинейной и зависеть от множества факторов [1]. Каждая эмоция имеет свое процентное соотношение в общем эмоциональном состоянии человека. Обозначим условно базовые эмоции, следующим образом: A (злость), H (радость), D (отвращение), F (страх), S (грусть), W (удивление), N (нейтральное состояние).

Тогда математическая модель, которая описывает проявление лжи по этим базовым эмоциям, может быть выражена следующим образом:

$$L = A * w_1 + H * w_2 + D * w_3 + F * w_4 + S * w_5 + W * w_6 + N * w_7,$$

где L – проявление лжи, $w_1, w_2, w_3, w_4, w_5, w_6, w_7$ – веса, которые определяют вклад каждой эмоции в выражение лжи. Эти веса могут быть определены через эмпирические исследования с помощью алгоритмов машинного обучения, например, метода случайного леса.

ОБУЧЕНИЕ АЛГОРИТМА СЛУЧАЙНОГО ЛЕСА

Для обучения алгоритма случайного леса с целью распознавания эмоций нами создан набор данных, состоящий из изображений, отображающих 7 различных эмоций конкретного человека. Датасет сбалансирован:

по 18 изображений на каждую эмоцию. В тренировочном датасете 14 изображений, а в тестовом – 4. На рис. 1 изображены фотографии из датасета, отображающие различные эмоции.



Рис. 1. Пример изображений из датасета для обучения по ключевым точкам лица

Для распознавания эмоций на микровыражениях лгущего человека и говорящего правду создан набор данных, состоящих из видео, где 12 видеоизображений – лживые ответы на вопросы, а 10 видеоизображений – правдивые. Далее вручную были отобраны изображения с микровыражениями лица, эмоции на которых отличаются от эмоций на кадрах ранее или после.

Распознавания эмоций по ключевым точкам лица осуществлялось с использованием алгоритма машинного обучения по методу случайного леса (RandomForestClassifier). Для извлечения ключевых точек лица использовалась Python библиотека MediaPipe [2].

Точность обучения составила 93%. Кроме точности, были рассмотрены и другие метрики анализа результатов классификации (табл. 1).

Таблица 1

Результаты классификации, отраженные в метриках: Precision, Recall, F1-score

Виды эмоций	Precision (точность)	Recall (полнота)	F1-score	Количество изображений
Злость	1,00	1,00	1,00	4
Отвращение	0,80	1,00	0,89	4
Страх	1,00	0,75	0,86	4
Радость	1,00	1,00	1,00	4
Нейтральность	0,80	1,00	0,89	4
Грусть	1,00	1,00	1,00	4
Удивление	1,00	0,75	0,86	4
Точность			0,93	28

Матрица ошибок обучения изображена на рис. 2. По матрице ошибок заметно, что одно изображение, относящееся к эмоции страха, ошибочно относится моделью к нейтральной эмоции. Такое возможно в случае, если эмоция страха небольшой интенсивности и внешне более похожа на нейтральное состояние. Так же изображение с эмоцией удивления оши-

бочно относится к эмоции отвращения. Возможно потому, что при отвращении могут быть широко раскрыты глаза, что, в основном, является ключевым признаком удивления.

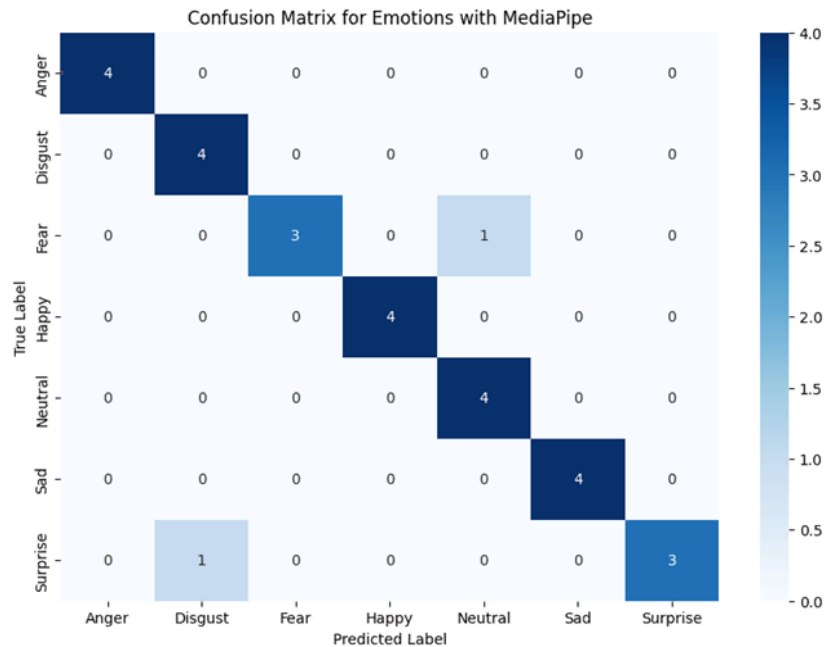


Рис. 2. Матрица ошибок

ПРОГНОЗИРОВАНИЕ ЭМОЦИОНАЛЬНОГО ФОНА ПО МИКРОВЫРАЖЕНИЯМ

По результатам обучения определены: минимальное, максимальное и среднее значения процентного соотношения для каждой эмоции.

Результаты анализа представлены в табл. 2.

Таблица 2

Статистика предсказанных эмоций моделью на основе ключевых точек лица

Эмоции/Значения	Минимальное, %	Максимальное, %	Среднее, %
Ложь			
Злость	29	67	47,9
Отвращение	14	26	18,4
Страх	0.0	2	1,7
Радость	4	15	8,3
Нейтральность	1	9	3,9
Грусть	0.0	2	0,4
Удивление	9	34	19,4
Правда			
Злость	12	24	15,3
Отвращение	15	18	17,1
Страх	1	2	1,86
Радость	6	15	8,57

Нейтральность	2	4	3,14
Грусть	0.0	1	0,43
Удивление	14	22	18,9

Из таблицы видно, что микровыражения на лице лгущего человека чаще всего являются следствием присутствия эмоций злости, отвращения и удивления, а во время произнесения правдивой информации эмоции имеют меньшую интенсивность, однако, среднее значение эмоции радости преобладает.

На рис. 3 изображены наиболее характерные результаты оценки эмоционального фона по фотографии лица, полученные по разработанной методике.

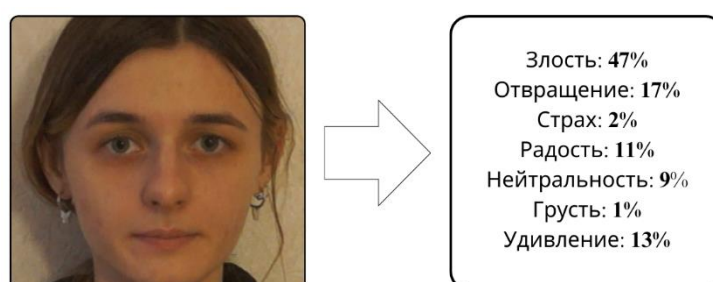


Рис. 3. Результаты предсказания эмоционального фона алгоритма случайного леса, обученного по ключевым точкам лица

ЗАКЛЮЧЕНИЕ

В ходе данной работы был рассмотрен алгоритм машинного обучения по методу случайного леса (RandomForestClassifier). Полученная точность, равная 93%, обуславливается правильным подбором гиперпараметров модели классификации, а также использованием сбалансированного, качественного датасета, собранного самостоятельно.

При анализе микровыражений, отобранных из кадров видео, было выявлено, что в видео, где человек лжет, содержится больше микровыражений, чем в видео, где говорит правду. Так же обнаружено, что интенсивность проявления эмоций выше у «лживых», чем у «правдивых» микровыражений.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Экман П. Психология лжи. Обмани меня, если сможешь. / Пол Экман. СПб.: Питер, 2021. С. 384.
2. MediaPipe Python tutorial [Электронный ресурс]. URL: <https://iqraanwar.medium.com/mediapipe-computer-vision-55821a144f7f> (дата обращения: 09.09.2023).

СИСТЕМА ДИСТАНЦИОННОГО УПРАВЛЕНИЯ И АНАЛИЗА ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ ФОТОПРИЕМНИКОВ И ФОТОЧУВСТВИТЕЛЬНЫХ МАТЕРИАЛОВ

**М.В. Лобанок, И.Д. Щербич, А.С. Семенович, М.А. Павлышко,
И.А. Романов**

*Белорусский государственный университет, Минск, Беларусь
E-mail: lobanokMV@bsu.by*

На базе монохроматора МДР-12 и микроконтроллера STM32 предложено аппаратное и программное решение для автоматизации измерений фотопроводимости и электролюминесценции. Предложена реализация дистанционного управления и анализа спектральных данных посредством протокола MQTT и микроконтроллера ESP32. Определена спектральная зависимость фототока гетероструктур SiC/Si.

Ключевые слова: микроконтроллеры; гетероструктуры; фототок; беспроводная связь.

Разработка и промышленное производство фоточувствительных многоспектральных приборных структур и ИК-фотоприемников требует проведения промежуточного контроля фотоэлектрических параметров, которые определяют характеристики конечных приборов [1]. Фотодетекторы, функционирующие в фотогальваническом режиме и отличающиеся низкими уровнями темновых токов, а также способностью к самостоятельной работе без внешнего электропитания, представляют особый интерес. Такие фотодетекторы особенно важны для систем, включающих в себя обширные массивы детекторов и использующих беспроводную оптическую связь [1,2].

В работе представлено аппаратно-программное решение для автоматизации системы на базе монохроматора МДР-12, которое позволяет исследовать спектральные зависимости фототока, а также электролюминесценцию. Одна из трудностей в разработке измерительного оборудования для параметров фотоприемников нового поколения заключается в недостатке унифицированных измерительных процедур, пригодных для внедрения в измерительные системы. Данную проблему можно решить, модифицируя существующие методы измерения для фотоприемников, описанные в стандарте ГОСТ 17772-88.

Монохроматор МДР-12 построен по схеме Черни-Тернера. В зависимости от типа измерения оптическая схема измерителя меняется. В случае измерения фотоотклика и фотопроводимости используется следующая схема. Источник света (вольфрамовая лампа СИ6-100, которая излучает пучок света в спектральном диапазоне 0,38-3,5 мкм) через конденсор освещает

щает входную щель монохроматора. Диспергированный свет параболическим объективом и поворотным зеркалом собирается в плоскости выходной щели, а затем попадает на исследуемый образец. Управление поворотным зеркалом происходит с помощью шагового двигателя, управляемого контроллером, а сигнал с образца проходит усиление на операционном усилителе и считывается через АЦП в контроллере STM32. Измерение других фотоэлектрических характеристик происходит по схеме спектрометра, представленной на рис. 1. В качестве детектора используется фотоумножитель ФЭУ-100.

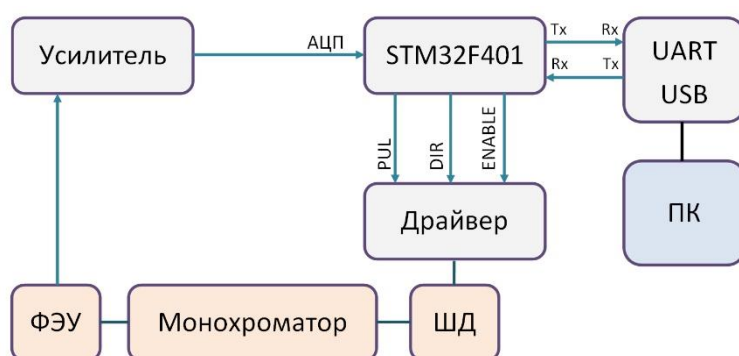


Рис. 1. Структурная схема спектрометра на основе малогабаритного монохроматора

Алгоритм работы спектрометра включает вращение дифракционной решетки шаговым двигателем с заданной скоростью во время сканирования спектра. Достижимый диапазон измерений составил 0,3-2,0 мкм с шагом 20 нм. Стоит отметить, что спектральное разрешение системы зависит от ширины выходной щели, скорости сканирования и фильтрации. Наличие богатой периферии микроконтроллера STM32F401 позволит в будущем реализовать систему цифровой фильтрации данных методом синхронного детектирования.

Для минимизации внешнего влияния на измерительную систему (засветка, тепловой шум, механические колебания) предложена реализация удаленного управления системой и анализа данных фотоэлектрических характеристик исследуемых структур с помощью аппаратно-программной платформы, представленной на рис. 2, на базе микроконтроллера ESP32 с использованием протокола прикладного уровня MQTT[3]. В используемую схему измерений (рис. 1) был добавлен микроконтроллер ESP32 для реализации беспроводного канала связи с измерительным стендом.

В качестве управляющего модуля выбрана процессорная система на кристалле ESP32, которая при малом энергопотреблении обладает достаточной вычислительной мощностью и богатством периферийных интерфейсов. Система поддерживает весь стек протоколов стандартов Wi-Fi

802.11n и Bluetooth 4.2, обеспечивая дополнительный функционал сопряжения периферийных датчиков через интерфейсы SPI/SDIO или I²C/UART. Кристалл системы включает в себя 32-разрядный процессор Tensilica Xtensa LX6 и контролеры Bluetooth BLE, Wi-Fi которые позволяют подключить разнообразные цифровые датчики (сенсоры) и легко организовать беспроводную связь с использованием доступных интернет протоколов. Для реализации режима синхронизации внутренних часов можно модернизировать протокол сообщения MQTT добавив метку времени отправки пакета для анализа задержки передачи, по аналогии с алгоритмом RTP протокола.

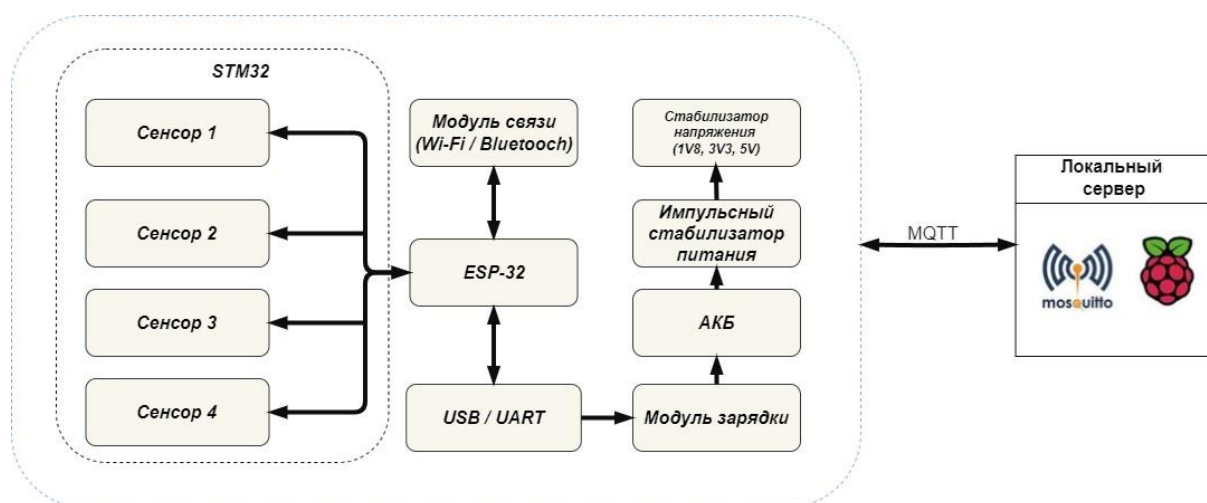


Рис. 2. Структурная схема аппаратной платформы

Для проведения функционального тестирования созданного макета было разработано тестовое программное обеспечение, которое состоит из двух частей: микропрограммы для узловых устройств и программного модуля для локального сервер – брокера MQTT сообщений.

Программные модули созданы на языке программирования «верхнего» уровня C++ с использованием программы-транслятора Visual Code. Для конфигурирования режимов работы модулей с датчиками и внешними устройствами использовались опубликованные на условиях лицензии свободного программного обеспечения (GPL - General Public License) библиотеки сообщества инженеров и радиолюбителей. Локальный Брокер – это открытая реализация Mosquitto Mqtt-Server который легко запускается на любом Raspberry устройстве. В нашем случае сервер был запущен на одноплатном компьютере Raspberry Pi 3. В качестве удалённой облачной системы была использована открытая платформа HiveMq с возможностью хранения безликих данных эксперимента (бесплатно в течении суток). Эта платформа сможет обеспечивать управление и передачу данных

по запросу от специализированного web-сервера для накопления и анализа статистики эксперимента с настроенными профилями [3].

В ходе работы был получен и подвергнут обработке в программе Origin массив экспериментальных спектров фототока для гетероструктур SiC/Si с необходимым разрешением и точностью. На рис. 3 приведена спектральная зависимость фототока гетероструктуры SiC/Si, полученной методом быстрой вакуумно-термической обработки кремния [1].

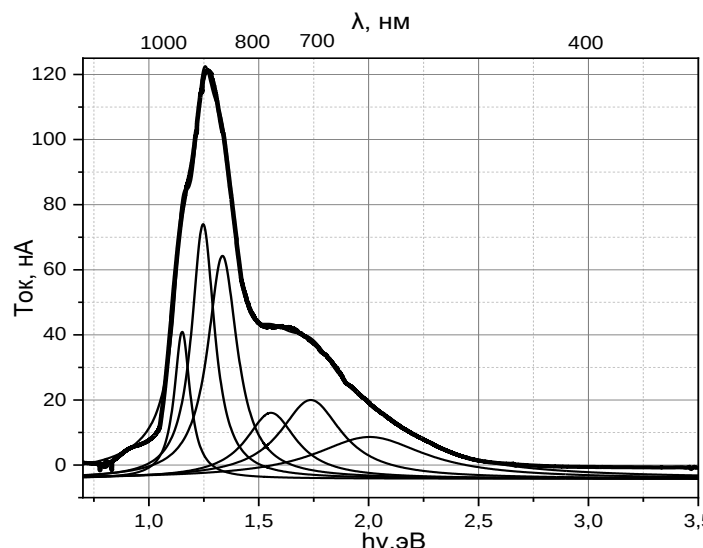


Рис. 3. Спектральная зависимость фототока гетероструктуры SiC/Si, полученная с помощью платформы Hivemq

Для определения влияния поглощения в слоях SiC и Si-подложке на спектре представлена аппроксимация функцией с несколькими лоренциалами. Гетероструктура SiC/Si генерирует фототок в диапазоне 650-1000 нм, а изменение генерируемого тока связано с влиянием кремниевой подложки и дефектов на границе раздела [1]. Из анализа следует вывод, что гетеропереход 3C-SiC/Si может быть использован в качестве многоспектрального фотодетектора

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. М.В. Лобанок, Н.В. Полонский, П.И. Гайдук. Фотоэлектрические характеристики гетероструктур SiC/Si. Журнал прикладной спектроскопии. 2023. №5 (90). С.775-779.
2. N. Youngblood, C. Chen, S. Koester, M. Li. Waveguide-integrated black phosphorus photodetector with high responsivity and low dark current. Nature Photon 2015; (9): С. 247-252.
3. Семенович А.С. Масштабируемый IoT тренажер спортсмена. Сборник научных работ студентов Республики Беларусь «НИРС 2020». С. 143-144.

АВТОМАТИЗАЦИЯ ИЗМЕРЕНИЙ ЭЛЕКТРОФИЗИЧЕСКИХ ХАРАКТЕРИСТИК ТОНКОПЛЕНОЧНЫХ МАТЕРИАЛОВ

М.В. Лобанок, В.А. Зайков, А.А. Гринкевич, И.Д. Щербич,
А.О. Суконкина

Белорусский государственный университет, Минск, Беларусь
E-mail: lobanokMV@bsu.by

На базе четырехзондового измерителя сопротивления ИУС-3 и микроконтроллера ESP32 предложено программное и схемотехническое решение для автоматизации измерений температурных электрофизических характеристик тонкопленочных материалов. В тестовом режиме на покрытиях TiAlCuN, нанесенных реактивным магнетронным распылением, проведено успешное испытание макета прибора.

Ключевые слова: микроконтроллеры; тонкие пленки; удельное сопротивление; неразрушающий контроль материалов.

За последние десятилетия в различных материалах, таких как оксиды и нитриды металлов, были обнаружены новые физические свойства, такие как термоэлектрический эффект, электронные фазовые переходы (переходы изолятор-металл и наоборот), зарядовое упорядочение и так далее [1]. Для характеристики этих материалов в тонкопленочном состоянии, важным параметром является удельное сопротивление (ρ) и его зависимость от температуры $\rho(t)$, поскольку это один из наиболее чувствительных индикаторов физических эффектов, вызванных изменением электронной зонной структуры. Для измерения сопротивления в литературе предложены различные методы и модели, среди которых четырехзондовый метод измерения, является наиболее востребованным [2].

Предложенное нами аппаратно-программное решение позволяет расширить возможности стандартного прибора ИУС-3, за счет увеличения диапазона измерений удельного сопротивления, реализации подогреваемого столика для калориметрических измерений, а также удаленного подключения и автоматической записи результатов измерений.

В общем случае, для контактных измерений электрических характеристик используется метод Ван дер Пау [2]. Однако, для таких измерений, когда толщина пленки много меньше расстояния между контактами, можно проводить более простое измерение слоевого (поверхностного) сопротивления (R_{cl}), не зависящее от линейных размеров. В этом случае удельное сопротивление рассчитывается из измеряемого слоевого сопротивления по следующей формуле:

$$\rho = \frac{R_{cl}}{d}, \quad (1)$$

где ρ – удельное сопротивление, R_{cl} – слоевое сопротивление, d – толщина покрытия (толщина покрытия d , в наших исследованиях измерялась методом растровой электронной микроскопии на сколе образца).

Способ измерения удельного сопротивления заключается в следующем: на плоскую поверхность исследуемого материала опускается измерительная голова из четырёх зондов, размещенных по прямой линии. Принципиальная схема измерения приведена на рис. 1.

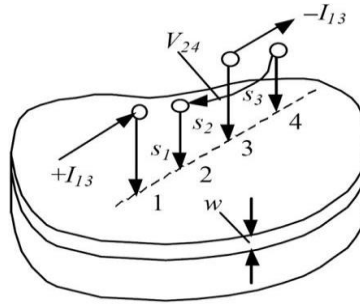


Рис. 1. Реализация четырехзондового метода

Через два внешних зонда пропускают ток I , а на двух внутренних зондах измеряют разность потенциалов V . Если расстояния между соседними зондами одинаковы и равны s , то удельное сопротивление ρ массивного образца рассчитывают по формуле:

$$\rho = \frac{V}{I} * 2\pi s, \quad (2)$$

В ходе дальнейших вычислений, с учетом разности потенциалов и геометрических особенностей образцов, а также поправочных коэффициентов получаем формулу для расчета слоевого сопротивления для стандартной контактной площадки измерителя ИУС-3:

$$R_{cl} = 4,53 \frac{V_{23}}{I_{14}}, \quad (3)$$

В общем случае точность измерения сопротивления четырехзондовым методом достаточно высока ($\sim 5\%$). Однако, трудности учёта различных факторов, таких, как локальный разогрев образца, давление зондов (в том числе и различия в давлении потенциометрических зондов), поверхностные и геометрические эффекты, приводят к уменьшению точности измерения ρ . Как показано в работе [3], погрешности измерений слоевого сопротивления могут достигать величины – 10%. Частичный учёт и уменьшение погрешности измерений достигается при измерении на серии одинаковых образцов или серии измерений на одном образце и последующем усреднении результатов, что и было выполнено в предложенном нами алгоритме измерений.

Экспериментально определено, что для обеспечения точности измерений в ИУС-3, ток, проходящий через пленку, должен измеряться с точностью до 1 мкА, а напряжение на центральных зондах измерительной головы – с точностью до 1 мкВ. Кроме того, минимальный шаг дискретизации значения напряжения должен быть 0,5 мкВ [4].

При реализации аппаратной части нами использовалась стандартная измерительная голова ИУС-3 и микроконтроллер семейства ESP32. Для контроля и поддержания температуры, изготовленного нами резистивного нагревательного столика, использовалась термопара К-типа, подключенная к 12-битному АЦП в модуле MAX 6675, а также в приборе для термостабилизации температуры был реализован ПИД-регулятор. Структурная схема измерителя представлена на рис. 2.

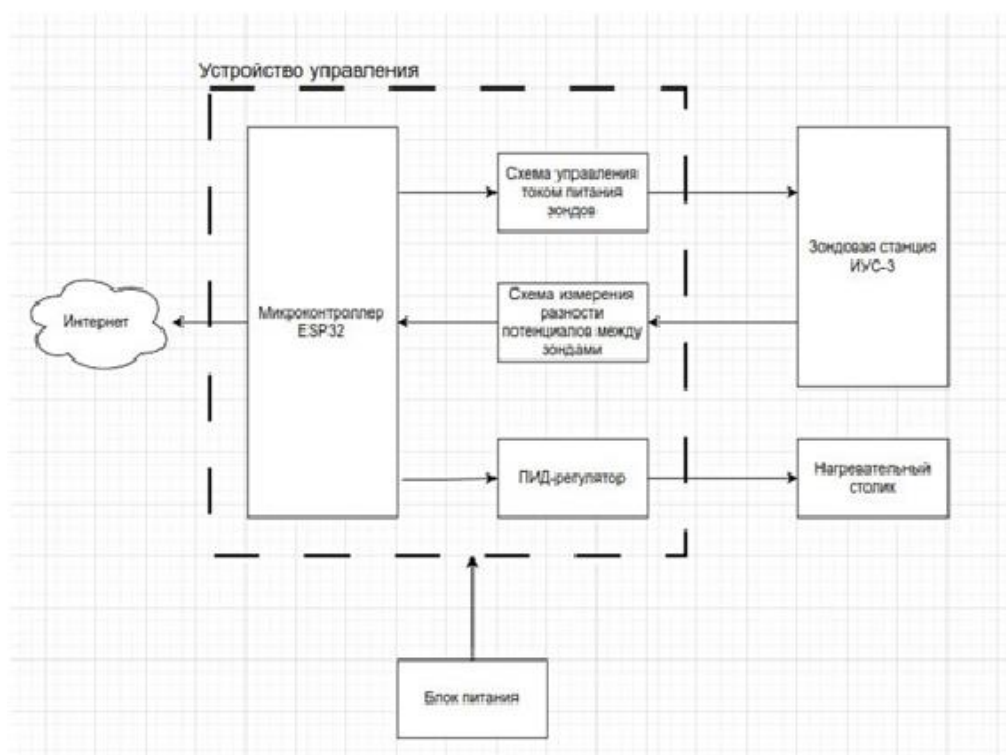


Рис. 2. Структурная схема измерителя

Алгоритм работы измерителя включает следующие процедуры: В начале измерения происходит запуск цикла подбора напряжения. Далее на исследуемый образец подается постоянное напряжение выбранного напряжения и измеряются ток и напряжение на центральных зондах измерительной головы ИУС-3. В случае, когда значение подаваемого напряжения достаточно мало (мкВ), то происходит увеличение напряжения на образце в 10 раз; Подобный цикл идет до тех пор, пока напряжение и ток на центральных зондах не станут равными некоторой существенной величине (напряжение от 0,1 В и ток не менее 2,0 мА). Следующим шагом, после

установления необходимого диапазона напряжения и тока, является проведение серии измерений и последующее усреднение. Итоговым результатом передаваемым на компьютер будет среднеарифметическое значение R_{cl} за N измерений.

На рис. 3 приведена зависимость удельного сопротивления покрытия TiAlCuN, полученного методом реактивного магнетронного распыления, от температуры. Треугольники – одна из серий измерений, а сплошная линия – результат усреднения по пяти сериям.

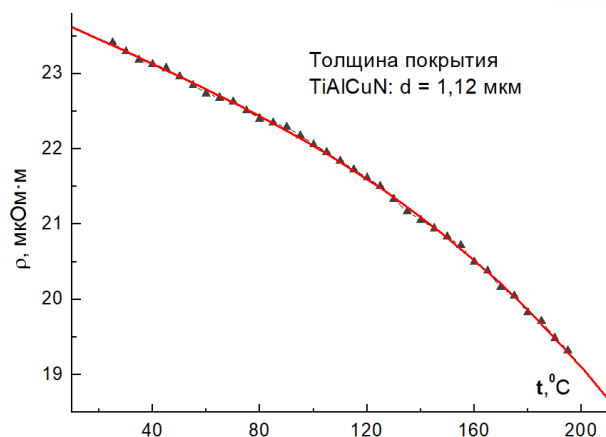


Рис. 3. Зависимость удельного сопротивления покрытия TiAlCuN от температуры

Из анализа зависимости удельного сопротивления покрытия TiAlCuN от температуры установлено, что температурный коэффициент сопротивления (ТКС) является отрицательной величиной, что свидетельствует о неметаллическом типе проводимости. Можно предположить наличие полупроводниковых свойств у тонкопленочного покрытия TiAlCuN [2].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. T. K. Roy, D. Sanyal, D. Bhowmick, A. Chakrabarti. Temperature dependent resistivity study on zinc oxide and the role of defects // Materials Science in Semiconductor Processing 16. 2013.. С. 332–336.
2. Determination of the temperature coefficient of resistance from micro four-point probe measurements / T. Marangoni, B. Guralnik, K. Borup [et al.] // Journal of Applied Physics 129, 165105. 2021.
3. М.В. Лобанок, Н.В. Полонский, С.В. Козодоев, П.И. Гайдук. Влияние термообработки на электрофизические характеристики контактов Pt к гетероструктурам SiC/Si. Известия Гомельского государственного университета имени Ф. Скорины. 2022. №6 (135). С.124-129.
4. Электрофизические характеристики структур TiAlCN/TiAlN, модифицированных наносекундной импульсной лазерной обработкой / О.С. Сиренко, С.А. Игнатович, И.Д. Щербич, М.В. Лобанок [и др.] // Квантовая электроника: Матер. XIV Международ. конф. Мн.: БГУ, 2023. С. 453-456

РЕАЛИЗАЦИЯ СТЕНДА ДЛЯ ИЗМЕРЕНИЙ СИГНАЛЬНЫХ ХАРАКТЕРИСТИК ЛАВИННОГО ФОТОДИОДА

М.С. Матусевич, С.С. Гринько, А.А. Скребец, М.В. Лобанок,
Е.В. Кадола

Белорусский государственный университет, Минск, Беларусь
E-mail: lobanokMV@bsu.by

Предложено программное и схемотехническое решение на базе микроконтроллера ESP32 для автоматизации измерений шумовых и сигнальных характеристик лавинного фотодиода. В тестовом режиме проведен анализ зависимости выходного сигнала фотодиода APD500um от нагрузочного сопротивления с целью увеличения коэффициента лавинного умножения.

Ключевые слова: микроконтроллеры; лавинный фотодиод; нагрузочное сопротивление.

Лавинные фотодиоды (ЛФД) являются одними из ключевых компонентов в современных оптических системах, применяемых в медицине и исследовательских целях. Способность лавинных фотодиодов к внутреннему усилению сигнала, высокие квантовая эффективность регистрации света (для Avalanche PhotoDiode более 50%) и временное разрешение (не хуже 1 нс) делает их незаменимыми для регистрации малых интенсивностей излучения [1]. Однако с увеличением усиления сигнала возрастает и уровень шума, что может снижать общую производительность прибора. Поэтому глубокое понимание шумовых и сигнальных характеристик ЛФД необходимо для оптимизации их работы и обеспечения высокой точности и надежности в различных приложениях. С другой стороны, автоматизация рутинных измерений выходного сигнала и вольт-амперных характеристик фотоприемников на основе перспективных гетероэпитаксиальных структур является важной задачей для развития оптоэлектроники.

Предложенное нами аппаратно-программное решение позволяет исследовать зависимость выходного сигнала от нагрузочного сопротивления детектора. При этом, применение микроконтроллера ESP32 позволило реализовать удаленное подключение и автоматическую запись измерений. Последующий анализ данных проводился с помощью MATLAB.

Принцип работы установки для изучения сигнальных и шумовых характеристик фотодетекторов основан на том, что оптическое излучение от He-Ne лазера мощностью 1мВт и длиной волны $\lambda = 0,63$ мкм, проходит через модулятор и регулируемую систему ослабителей, а затем попадает на фотодиод. Ослабленный до необходимого уровня световой поток направляется на исследуемый фотоприемник и фокусируется на его рабо-

чей площадке с помощью объектива. Электрический режим работы фотоприемника устанавливается с помощью регулируемого источника напряжения. Режим счета фотонов реализуется по схеме пассивного гашения лавины [2], когда ЛФД включается последовательно с токоограничивающим резистором R . В качестве нагрузочного резистора включается модуль программируемого магазина сопротивлений для измерения зависимости сигнала от нагрузочного сопротивления. От источника питания на лавинный фотодиод подается напряжение питания близкое или превышающее напряжение пробоя р–п-перехода. Если при таком напряжении в область микроплазменного пробоя р–п-перехода попадает фото- или термогенерированный носитель заряда, то он инициирует лавинный пробой.

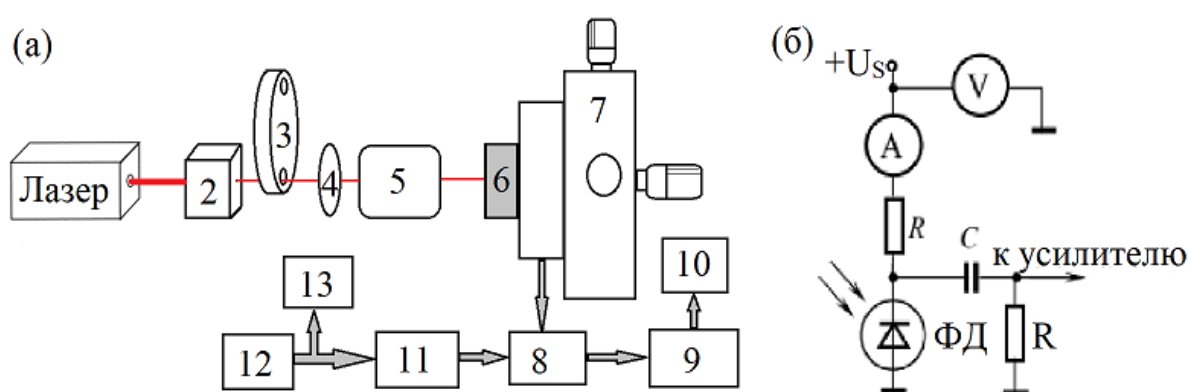


Рис. 1. Структурная схема установки для исследования сигнальных характеристик полупроводниковых фотоприемников (а) и схема включения фотодиода (б):

1 – лазер; 2 – система оптических ослабителей; 3 – диск механического модулятора; 4 – фокусирующая линза; 5 – микроскоп; 6 – исследуемый фотоприемник; 7 – трехкоординатный оптический столик; 8 – электрическая цепь фотоприемника; 9 – измерительный усилитель; 10 – осциллограф; 11 – амперметр; 12 – источник питания фотоприемника; 13 – вольтметр

Оптическое излучение преобразуется фотоприемником в электрический сигнал, который поступает на вход измерительного усилителя. В качестве последнего используется усилитель селективный У2-8. Контроль формы сигнала, с выхода усилителя, осуществляется с помощью осциллографа.

В данном случае чувствительность ЛФД P работающего в фотодиодном режиме рассчитывается из ВАХ фотодиода:

$$I = I_d \left(\exp \left[\frac{e(U_R - U_S)}{kT} \right] - 1 \right) - I_{ph}, \quad (1)$$

где U_S – напряжение источника питания, U_R – падение напряжения на сопротивлении нагрузки $U_R = IR$, I_{ph} – фототок ЛФД, I_d – темновой ток, роль которого в данном случае играет обратный ток р–п-перехода.

Если абсолютная величина разности $U_S - U_R \gg kT/e$, а фототок выражается через мощность падающего излучения и токовую чувствительность фотодиода $I_{ph} = SP$, то напряжение на нагрузке фотодиода будет:

$$U_R = (I_d + SP)R. \quad (2)$$

Из формулы (2) можно определить дифференциальную вольттовую чувствительность по следующей формуле:

$$(S_U)_U = (I_d + SP)R. \quad (3)$$

Переключение между нагрузочными резисторами в модуле программируемого магазина сопротивлений реализовано по классической схеме декадного магазина с помощью герметичных реле, управляемых через микроконтроллер. Измерение выходного сигнала ЛФД происходит через 24-битное АЦП, затем полученный массив данных отправляются на компьютер для дальнейшей обработки. Чтобы подключить все реле к микроконтроллеру ESP используются дополнительные контроллеры Atmel 328, к которым подключаются реле. В свою очередь, внешнее АЦП подключено напрямую к ESP32 для лучшего быстродействия. Такая связка операционного усилителя и АЦП для измерения спектральной зависимости фототока гетероструктур (порядка десятка нА) [3]. Помимо реле, к контроллеру Atmel 328 можно подключить модуль MAX 6675 с термпарой, а также транзистор с элементом Пельте для термостабилизации фотодиода. Структурная схема измерителя представлена на рис. 2. Механический модулятор сигнала совместно с оптическим датчиком, подключенным к контроллеру, позволяет реализовать синхродетектирование, что позволяет повысить точность измерений и получить данные о фотоотклике приборной структуры [3].



Рис. 2. Структурная схема измерителя

Алгоритм работы измерителя включает следующие процедуры: Вначале при отключенном питании лазера производятся измерения темнового тока. Далее включается лазер и производится серия измерений

выходного сигнала фотодиода с поочередным переключением нагрузочного сопротивления.

Итоговым результатом является передаваемые на компьютер среднеарифметическое значение U_R и I_{ph} за N измерений, которые в дальнейшем можно использовать для расчета чувствительности по формуле (3). На рис. 3 приведена зависимость выходного сигнала лавинного фотодиода APD500um от нагрузочного сопротивления.

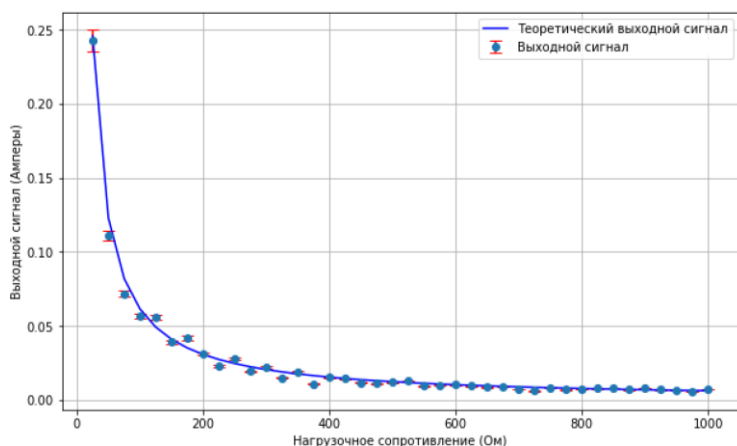


Рис. 3. Зависимость выходного сигнала лавинного фотодиода от нагрузочного сопротивления

Из приведенного графика следует, что при увеличении сопротивления нагрузки, выходной сигнал фотодиода (фототок) нелинейно уменьшается, что говорит о снижении коэффициента лавинного умножения. Вероятно, это связано увеличением времени зарядки внутренней емкости фотодиода и тем самым снижением скорости реакции на изменение светового сигнала.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. И.Р. Гулаков, А.О. Зеневич, О.В.Кочергина, Т.А. Матковская. Характеристики германиевых лавинных фотодиодов в режиме счета фотонов. Известия Национальной академии наук Беларуси. Серия физико-технических наук. 2022. Т. 67, №2. С. 222–229.
2. Многоканальная квантовая система волоконно-оптической безопасной связи. / О.К. Барановский [и др.] // Укр. науч. журн. информ. технологий. 2016. Т. 22, №2. С. 156–162.
3. М.В. Лобанок, Н.В. Полонский, П.И. Гайдук. Фотоэлектрические характеристики гетероструктур SiC/Si. Журнал прикладной спектроскопии. 2023. №5 (90). С.775–779.

TESTING AND ANALYZING THE PERFORMANCE OF SINGLE-BOARD COMPUTERS FOR AI TASKS

D. A. Paulenka, E. V. Snezhko, V. A. Kovalev

*The United Institute of Informatics Problems of the
National Academy of Sciences of Belarus, Minsk, Belarus
E-mail: dmitri.pavlenko@gmail.com,
eduard.snezhko@gmail.com, vassili.kovalev@gmail.com*

Performance tests were conducted on the Raspberry Pi 4 Model B (Pi4) and Cool Pi 4 Model B (CoolPi) single board computers, as well as the Google Coral USB Accelerator with Google Edge TPU (Coral). Their performance for AI tasks was analyzed. CoolPi handles classification using MobileNet v3 well, but is 4 times slower than similar computations on Coral. Classification on Pi4 is 22 times slower than Coral. CoolPi outperforms Pi4 by a factor of 2 to 3 in data copying and compression tasks and almost 6 times in neural network computations. Coral gives a significant acceleration of AI tasks, but there are features related to the novelty of the technology.

Keywords: single-board computers; neural processors; neural network accelerators; AI accelerators; performance testing; benchmarking.

ТЕСТИРОВАНИЕ И АНАЛИЗ ПРОИЗВОДИТЕЛЬНОСТИ ОДНОПЛАТНЫХ КОМПЬЮТЕРОВ ДЛЯ РЕШЕНИЯ ЗАДАЧ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Д. А. Павленко, Э. В. Снежко, В. А. Ковалев

*Объединенный институт проблем информатики Национальной академии наук
Беларуси, Минск, Беларусь
E-mail: dmitri.pavlenko@gmail.com,
eduard.snezhko@gmail.com, vassili.kovalev@gmail.com*

Проведены тесты производительности одноплатных компьютеров Raspberry Pi 4 Model B (Pi4) и CoolPi 4 Model B (CoolPi), а также USB-ускорителя Google Coral с TPU Google Edge (Coral). Проанализирована их производительность в задачах искусственного интеллекта. CoolPi хорошо справляется с классификацией с помощью MobileNet v3, но в 4 раза медленнее, чем аналогичные вычисления на Coral. Классификация на Pi4 выполняется в 22 раза медленнее, чем на Coral. CoolPi превосходит Pi4 в 2-3 раза в задачах копирования и сжатия данных и почти в 6 раз в нейросетевых вычислениях. Coral дает значительное ускорение выполнения задач ИИ, но есть и особенности, связанные с новизной технологии.

Ключевые слова: одноплатные компьютеры; нейронные процессоры; ускорители нейронных сетей; ИИ-ускорители; тестирование производительности; бенчмаркинг.

INTRODUCTION

Modern image processing methods and algorithms must utilize neural network-based artificial intelligence (AI) techniques to be innovative and competitive. The rapidly growing market for devices that use specialized microprocessors for efficient neural network training and/or operation requires new ways to practically implement neural network methods on these devices.

Two single-board computers (SBC) and one neural processor were selected and purchased for the performance tests:

- SBC Raspberry Pi 4 Model B (hereinafter referred to as Pi4) [1];
- SBC Cool Pi 4 Model B (hereinafter referred to as CoolPi) [2];
- Google Coral USB Accelerator with Google Edge TPU (hereinafter referred to as Coral) [3], which is not a standalone computer.

PERFORMANCE TESTING AND ANALYSIS

Figures 1-3 show the results of the performance tests for the three small devices (Pi4, CoolPi, Coral) and the Desktop personal computer (PC) with chipset Intel Lynx Point Z87 and CPU Intel i7-4770.

Desktop tests are conducted for illustrative and comparative purposes. It was not possible to find free cross-platform software for Windows and Linux, so the Docker containers used for Windows were "phoronix/pts" [4] and "ubuntu" [5].

In Figure 1, the smaller the bar on the chart, the better. For clarity, Figure 1 shows the time scale on a logarithmic scale.

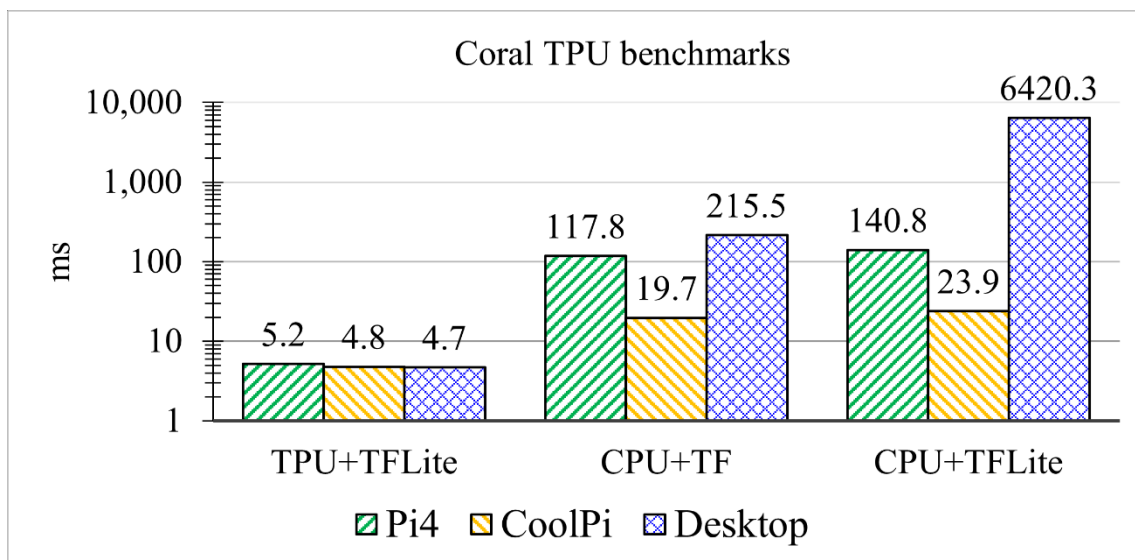


Fig. 1. Performance testing of the Coral TPU

"TPU+TFLite" contains performance test results for TPU Coral and TensorFlow Lite (TFLite) library. "CPU+TF" contains results for CPU and TensorFlow (TF) library. "CPU+TFLite" contains results for CPU and TensorFlow Lite (TFLite) library. TPU can only work with the TFLite library, so the fourth combination "TPU+TF" is not applicable.

Performance tests for Figure 1 were conducted on the classification task. The MobileNet v3 model pre-trained on ImageNet images was used for classification. The model files can be found in the Coral repository [6]:

- file "tf2_mobilenet_v3_edgetpu_1.0_224_ptq_edgetpu.tflite" (TPU);
- file "tf2_mobilenet_v3_edgetpu_1.0_224_ptq.tflite" (CPU).

The performance tests software can be found in the files "tf_lite_benchmarks.py" and "tf_lite.py" [7]. The log file with the test results is located in the GitHub repository file "data/coral-tpu-benchmark-results.txt" [7].

Figure 1 shows that the speed of calculations on Coral TPU ("TPU+TFLite") differs insignificantly for different devices and significantly outperforms CPU calculations. A comparison of the "CPU+TF" and "CPU+TFLite" showed that TensorFlow was slightly faster than TensorFlow Lite, except for the Desktop.

Perhaps, for Desktop PC, the significant 30-fold speed difference of "CPU+TF" (215.5 ms) and "CPU+TFLite" (6,420.3 ms) is due to some internal optimizations of the TF and TFLite libraries.

In Figure 2, the higher the bar of the chart, the better.

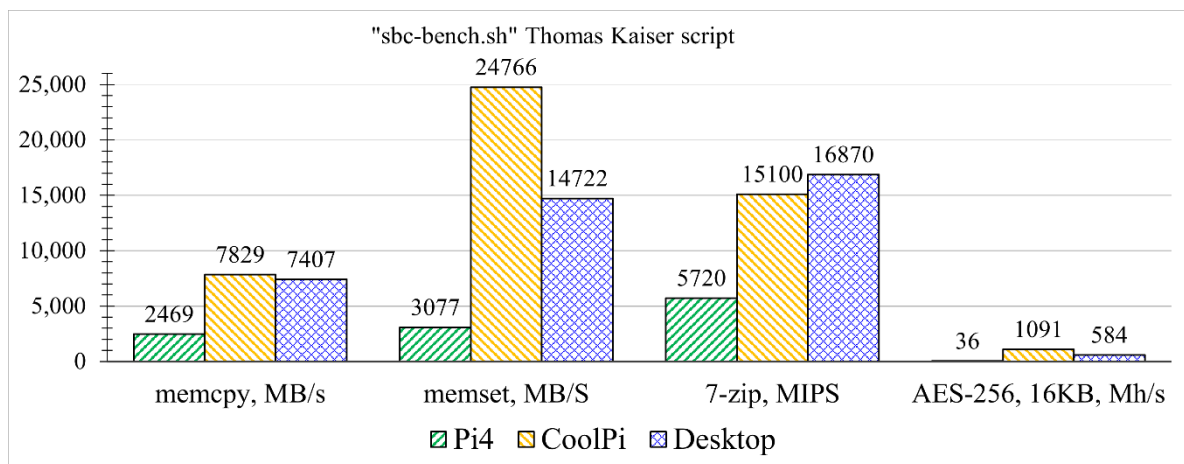


Fig. 2. Performance testing with the "sbc-bench.sh" script

In Figure 2, performance testing is performed using Thomas Kaiser software «sbc-bench.sh» [8], which is designed specifically for SBC performance tests.

"memcpy" is the speed of copying data from one block of memory to another using the "memcpy" function of the C/C++ programming language.

"memset" is the memory block initialization speed using the C/C++ "memset" function. "7-zip" is the speed of data compression using "7-zip" software. "AES-256" is the data encryption speed using "AES-256" algorithm. Performance tests have shown that only the Pi4 has no throttling, perhaps, due to the poor cooling system of CoolPi and Desktop PC.

The log file with the test results is located in the GitHub repository file "data/Thomas-Kaiser-sbc-bench-results.txt" [7].

In Figure 3, the higher the bar of the chart, the better.

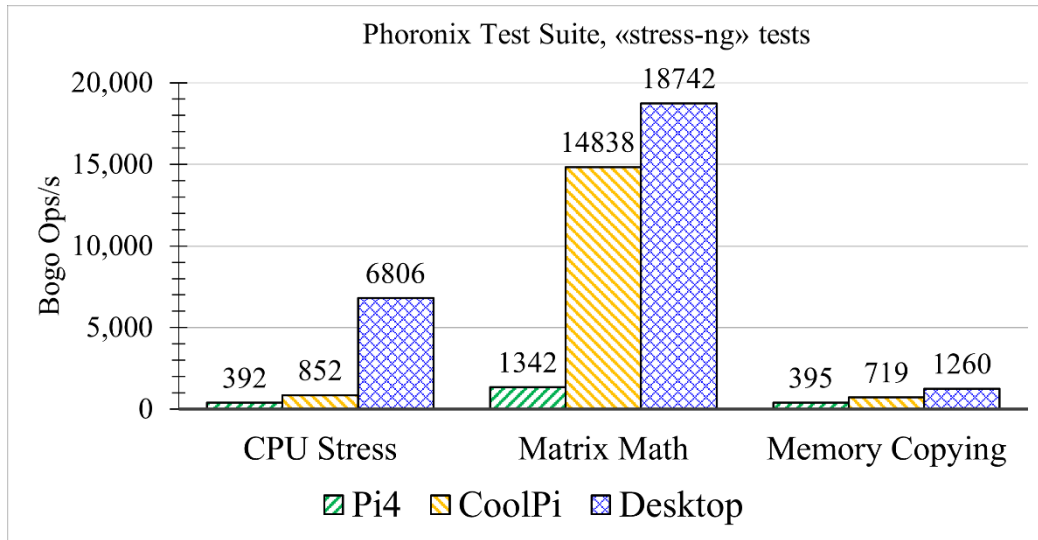


Fig. 3. Performance testing with Phoronix Test Suite, "stress-ng" test

In the Phoronix Test Suite software [9], performance is measured in Bogo Ops/s (bogus operations per second), which is a way of measuring the speed at which a computer executes instructions in the Linux kernel.

Phoronix Test Suite software [9] has hundreds of different performance tests, from which the "stress-ng" test was selected, and performance tests have been carried out: "CPU Stress", "Matrix Math" and "Memory Copying".

The log file with the test results is located in the GitHub repository file "data/phoronix-stress-ng-results.txt" [7].

CONCLUSIONS

Coral TPU gives a significant acceleration of AI tasks, but there are features related to the novelty of the technology that will be investigated in future works. CoolPi is 4 times slower than Coral. Pi4 is 22 times slower than Coral. CoolPi outperforms Pi4 by a factor of 2 to 3 in data copying and compression tasks and almost 6 times in AI tasks.

ACKNOWLEDGEMENTS

This work was carried out with the financial support of the Complex-SG space program.

REFERENCES

1. Raspberry Pi 4. [Electronic resource]. URL: <https://www.raspberrypi.com/products/raspberry-pi-4-model-b> (date of access: 27.03.2024).
2. Cool Pi 4 Model B – A much faster alternative to Raspberry Pi 4 SBC [Electronic resource]. URL: <https://www.cnx-software.com/2022/12/04/cool-pi-4-model-b-powerful-raspberry-pi-4-alternative> (date of access: 27.03.2024).
3. USB Accelerator [Electronic resource]. URL: <https://coral.ai/products/accelerator> (date of access: 27.03.2024).
4. phoronix/pts [Electronic resource]. URL: <https://hub.docker.com/r/phoronix/pts> (date of access: 27.03.2024).
5. ubuntu [Electronic resource]. URL: https://hub.docker.com/_/ubuntu (date of access: 27.03.2024).
6. google-coral/test_data [Electronic resource]. URL: https://github.com/google-coral/test_data. (date of access: 27.03.2024).
7. foobar167/junkyard/coral_tpu [Electronic resource]. URL: https://github.com/foobar167/junkyard/tree/master/coral_tpu (date of access: 27.03.2024).
8. ThomasKaiser/sbc-bench [Electronic resource]. URL: <https://github.com/ThomasKaiser/sbc-bench> (date of access: 27.03.2024).
9. Phoronix Test Suite [Electronic resource]. URL: <https://www.phoronix-test-suite.com>. (date of access: 27.03.2024).

ДЕМОДУЛЯЦИЯ И ДЕКОДИРОВАНИЕ QPSK-РАДИОСИГНАЛА СТАНДАРТА CCSDS

М.А. Павлышко, С.Н. Семенович, В.А. Прохоров

Белорусский государственный университет, Минск, Беларусь

E-mail: pauyshka@bsu.by

Рассмотрены стандарты передачи данных CCSDS. Предложен граф для демодуляции QPSK-радиосигнала в GNU Radio. Предложено программное обеспечение для быстрого декодирования и анализа декодированного битового потока данных. Разработанные граф и ПО протестированы на записи QPSK-радиосигнала, переданного космическим аппаратом.

Ключевые слова: CCSDS; радиосвязь; QPSK; GNU Radio.

В настоящее время актуальна проблема быстрой и надежной передачи данных с космических аппаратов. Для оптимизации затрат на разработку и эксплуатацию систем приема данных ключевую роль играет применение установленных международных стандартов, таких как рекомендации «Консультативного комитета по космическим системам передачи данных» (CCSDS). Стандарты CCSDS обеспечивают необходимую унификацию технических процессов и интерфейсов, что позволяет сократить время и ресурсы, необходимые для интеграции различных компонентов систем передачи данных.

В частности, комитетом был разработан комплекс рекомендаций для обеспечения эффективной передачи телеметрической информации между космическими аппаратами и наземными станциями. Рекомендации включают в себя протоколы кодирования, форматы данных и методы синхронизации, которые способствуют надежной коммуникации в условиях открытого космоса [1].

Для модуляции битового потока данных, согласно рекомендации CCSDS, может применяться четырехпозиционная фазовая модуляция (QPSK). Для защиты данных от шумов физического канала битовый поток кодируется одним или двумя следующими кодами: блочный код Рида-Соломона (255, 233) или же сверточный код (7, 12) [2]. Для обеспечения обнаружения кодовых слов в передаваемом потоке данных на приемной стороне и их правильного декодирования при формировании кадра для передачи по физическому каналу применяется маркер синхронизации (ASM). Маркер синхронизации является постоянным, имеет длину 4 байта, добавляется в начало каждого кодового слова и принимает значение *1ACFFC1D* [3].

Для корректной демодуляции сигнала, принимающая система требует, чтобы входящий сигнал имел достаточную плотность битовых переходов для обеспечения синхронизации декодера. Входящий сигнал также

не должен содержать значительных спектральных линий и паттернов, мешающих синхронизации и проверке кодовых слов. Для достижения этих требований *CCSDS* рекомендует применение псевдослучайного рандомизатора [4]. Псевдорандомизатор обычно представляет собой регистр сдвига с линейной обратной связью (*LFSR*), полином которого задается следующей формулой:

$$h(x) = x^8 + x^7 + x^5 + x^3 + 1 \quad (1)$$

В рамках проведенных экспериментов была проведена демодуляция и декодирование сигнала с космического аппарата с символьной скоростью 14,336 Мсимв/с и кодированием *QPSK*. Для синхронизации принимаемого потока и снятия сверточного кодирования алгоритмом Витерби использовался разработанный граф для ПО *GNU Radio* (рис. 1). Для синхронизации символов применен метод максимального правдоподобия.

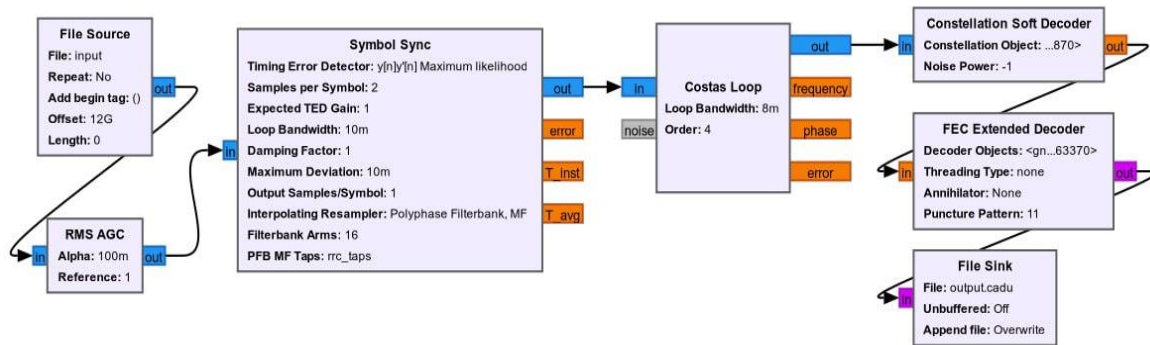


Рис. 1. Граф демодуляции и декодирования *QPSK*-радиосигнала для ПО *GNU Radio*

Для дальнейшей обработки принятого битового потока было разработано ПО на языке программирования *C++*. Конфигурация ПО происходит посредством соединения функциональных модулей в граф (рис. 2).

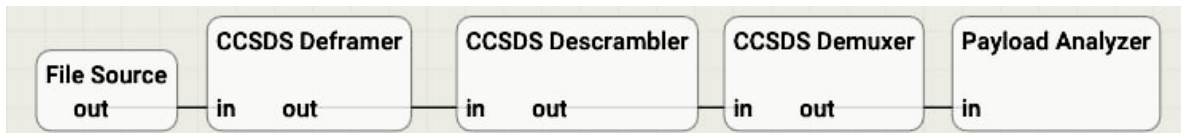


Рис. 2. Граф анализа сигнала в разработанном ПО

Модуль *CCSDS Deframer* обеспечивает синхронизацию битового потока и выделение информационного кадра посредством поиска в битовом потоке метки синхронизации (*ASM*) и последующим копированием заданного количества бит после нее. Блок *CCSDS Descrambler* выполняет сня-

тие псевдорандомизации с потока данных. Для оптимизации скорости выполнения вместо использования сдвигового регистра выполняется операция исключающего ИЛИ между битовым потоком и заранее рассчитанным массивом. Блок *CCSDS Demuxer* вычисляет контрольную сумму *CRC16* принятого кадра и сверяет ее с суммой, которая хранится в конце кадра. Также этот блок удаляет из дальнейшей обработки кадры, которые содержат специальный паттерн для заполнения радиоканала в случае отсутствия данных для передачи. Блок *Payload Analyzer* выполняет дальнейший разбор кадров в зависимости от конкретного космического аппарата (КА).

Разработанное ПО было успешно протестировано на сигнале КА с символьной скоростью 14,336 Мсимв/с, модуляцией QPSK и сверточным кодированием (рис. 3). Коэффициент битовых ошибок (BER) составил $<10^{-4}$.

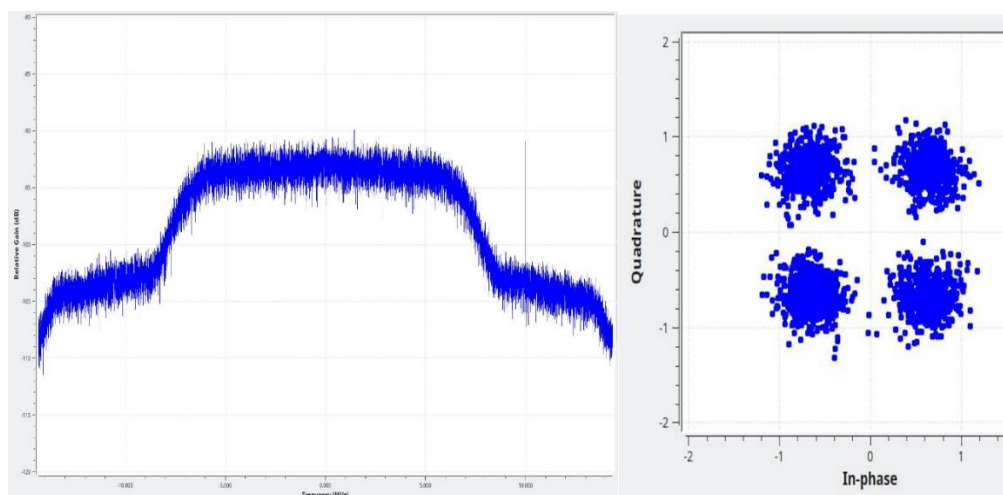


Рис. 3. Результат демодуляции принятого от КА сигнала разработанным ПО

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. CCSDS 130.3-G-0.2: CCSDS Draft Green Book. // Iss. 0.2. Sagamihara, Japan, 2004.
2. Минервин А. А., Поляков А. Ю. ИСПОЛЬЗОВАНИЕ ПРОТОКОЛОВ CCSDS В СОВРЕМЕННОЙ ТЕЛЕМЕТРИЧЕСКОЙ СИСТЕМЕ // Международный журнал экспериментального образования. 2010. № 3. С. 50;
3. Р.П. Богуш, И. Ю. Захарова, В.М. Чертков, Н.М. Наумович. Моделирование алгоритмов обработки данных радара с синтезированной апертурой космического аппарата для их передачи и формирования радиолокационного изображения // Вестник Полоцкого государственного университета. Серия С, Фундаментальные науки. 2018. № 12. С. 2–8.
4. CCSDS 131.0-B-5: TM Synchronization and Channel Coding // Iss. 5. Washington DC., USA, 2023.

СТЕНД ЭМУЛЯЦИИ ADS-B РАДИОСИГНАЛОВ ДЛЯ ТЕСТИРОВАНИЯ МОДУЛЯ СМКА BSUSAT-2

М.А. Павлышко, С.Н. Семенович, А.С. Семенович, М.А. Суровцев

Белорусский государственный университет, Минск, Беларусь

E-mail: pauyshka@bsu.by

Рассмотрено применение протокола ADS-B. Рассмотрена структура кадра ADS-B и особенности генерации радиосигнала ADS-B. Предложена реализация стенда для генерации и передачи данных в формате ADS-B. Протестирована реализация генератора пакетов ADS-B посредством приёма сгенерированных пакетов модулем наноспутника BSUSAT-2. Приведены результаты тестирования разработанного генератора сигналов.

Ключевые слова: ADS-B; гражданская авиация; наноспутник; SDR; BSUSAT-2.

Начиная с 1970-х годов, наблюдается значительный рост объема авиаперевозок, что увеличивает нагрузку на инфраструктуру и системы управления воздушным движением, требуя от них поддержания баланса между разнесением воздушных судов в пространстве и обеспечением высокой пропускной способности. Для решения задачи были разработаны новые технологические решения, среди которых выделяется система ADS-B.

Целью внедрения системы ADS-B является повышение безопасности полетов за счет добавления дополнительного уровня наблюдения и коммуникационных возможностей между воздушными судами и системами управления воздушным движением. В настоящее время использование ADS-B стало обязательным во многих частях мира, где законодательство требует его применения на борту воздушных судов, выполняющих полеты в наиболее загруженных секторах воздушного пространства, особенно в районах крупных международных аэропортов. Сигналы ADS-B содержат важную информацию, такую как идентификационный номер судна, его текущее местоположение, скорость, высоту полета и другие критически важные параметры [1].

В рамках создания наноспутника BSUSAT-2 был разработан модуль полезной нагрузки с приемником сигналов ADS-B. Основной целью модуля полезной нагрузки с приемником сигналов ADS-B является обеспечение получения точных и своевременных данных о положении и движении воздушных судов в зонах, где традиционные наземные станции не способны обеспечить адекватное покрытие, включая морские пространства и горные массивы. Модуль включает в себя приемник ADS-B, кото-

рый способен в реальном времени принимать сигналы, передаваемые воздушными судами, и микроконтроллер для последующей обработки данных.

На уровне физической среды ADS-B работает на двух радиочастотах: 1030 МГц для активного опроса, например, с вышек УВД, радаров или других воздушных судов, и 1090 МГц для активного ответа или широковещательных сообщений. Поскольку тестируемый модуль наноспутника BSUSAT-2 поддерживает только прием ответов и широковещательных сообщений на частоте 1090 МГц, реализация обмена на частоте 1030 МГц не проводилась.

На уровне линии передачи данных сообщение ADS-B инкапсулируется в кадры Mode-S [2]. Для передачи данных посредством Mode-S используется импульсно-позиционная модуляция (PPM). В этой схеме модуляции биты данных кодируются импульсами, каждый из которых имеет длительность 0,5 мкс и разное время начала относительно синхронизирующего импульса. Таким образом, скорость передачи данных для стандарта ADS-B фиксирована и составляет 1 Мбит/с (рис 1).

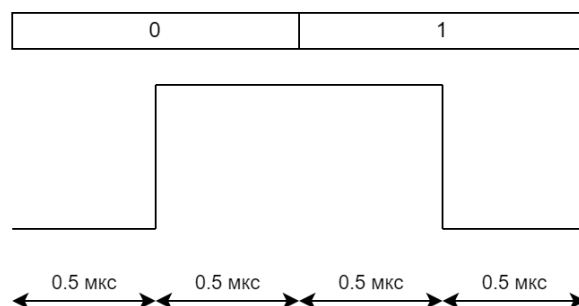


Рис. 1. Импульсно-позиционная модуляция для кадров Mode-S

Кадры ADS-B состоят из преамбулы и блока данных. Преамбула длится 8 мкс и представляет собой четыре импульса, длительностью 0,5 мкс с интервалами до первого 1, 3,5 и 4,5 мкс соответственно. Преамбула используется для синхронизации приемника ADS-B к принимаемому потоку. Блоки данных имеют длину 56 или 112 бит и используются для кодирования различных сообщений согласно формату нисходящей линии связи (DF, рис 2). Для симуляции широковещательных ответов от судов гражданской авиации необходимо использовать DF=17.

Преамбула 16 бит	DF 5 бит	CA 3 бита	ICAO 24 бита	ME 56 бит	PI 24 бита

Рис. 2. Формат кадра широковещательного ответа ADS-B:

DF – формат данных, CA – режим работы транспондера, ICAO – идентификационный номер судна, ME – полезная нагрузка, PI – контрольная сумма

В первых пяти битах поля ME хранится код типа сообщения (ТС, Type Code), который определяет формат оставшегося содержимого поля ME. Хранящееся внутри пакета сообщение может иметь один из следующих типов:

- 1–4 – идентификатор воздушного судна (позывной);
- 5–8 – координаты на поверхности;
- 9–18 – координаты в воздухе (барометрический высотомер)
- 19 – скорость в воздухе;
- 20–22 – координаты в воздухе (высотомер с применением ГНСС);
- 23–27 – резерв;
- 28–31 – состояние воздушного судна.

Для тестирования проверки работоспособности модуля полезной нагрузки наноспутника был разработан стенд для генерации сигнала и его передачи (рис. 3).

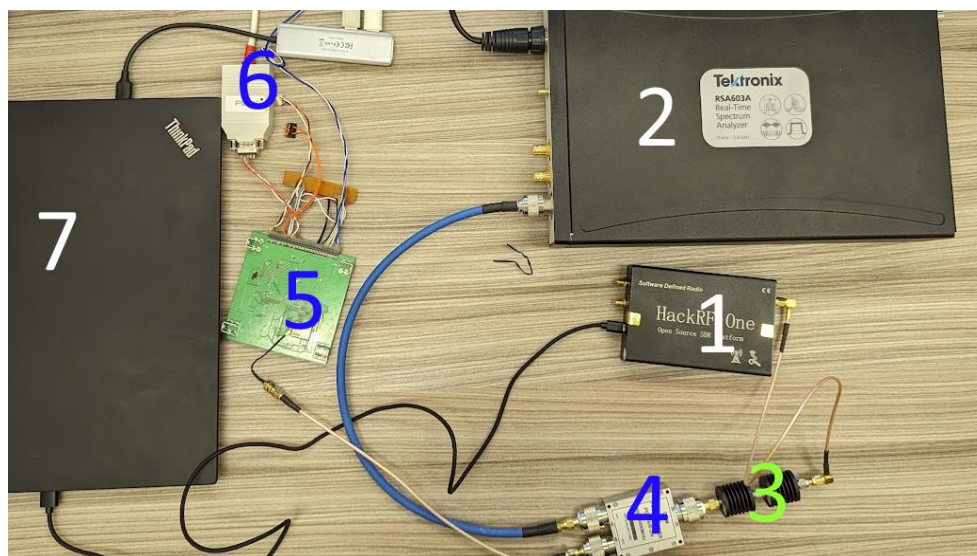


Рис. 3. Стенд для передачи и приема широковещательных ADS-B пакетов:

1 – SDR-передатчик HackRF One, 2 – анализатор спектра Tektronix RSA603A, 3 – аттенюатор, 4 – делитель мощности, 5 – модуль полезной нагрузки наноспутника BSUSAT-2, 6 – USB-CAN адаптер для связи с модулем полезной нагрузки, 7 – ПК для управления стендом

Для формирования пакетов данных в формате ADS-B были разработаны программные инструменты на языке Python, и с использованием GNU Radio

осуществлена их передача. Для проверки всех функциональных возможностей модуля полезной нагрузки наноспутника были сгенерированы ADS-B пакеты типов 4 (позывной), 11 (координаты) и 19 (скорость) для определённого количества случайных воздушных судов. Все сгенерированные пакеты были успешно приняты модулем СМКА BSUSAT-2 (см. рис. 4).

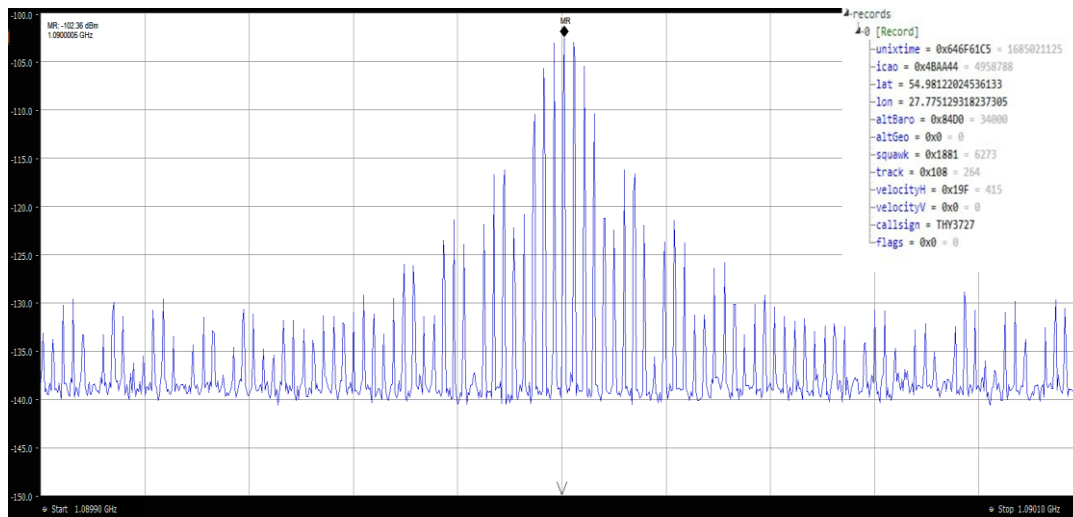


Рис. 4. Спектр передаваемого сообщения (размах – 200 кГц) и результат его приема и обработки модулем полезной нагрузки наноспутника BSUSAT-2

ВЫВОДЫ

Разработанный стенд для тестирования приемников стандарта ADS-B представляет собой комплексное решение для проверки и отладки устройств, принимающих сигналы в формате ADS-B. Стенд позволяет имитировать работу воздушных судов и генерировать информацию о позывном, координатах и скорости. Использование стенда обеспечивает возможность оценки надёжности и корректности работы приёмных устройств в различных условиях и сценариях использования.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. A. Costin, A. Francillion. Ghost in the Air (Traffic): On insecurity of ADS-B protocol and practical attacks on ADS-B devices // BLACKHAT 2012, Las Vegas, NV, USA.
2. J. Sun. The 1090 Megahertz Riddle: A Guide to Decoding Mode S and ADS-B Signals // TU Delft OPEN Publishing, 2021. P. 160. DOI: 10.34641/mg.11
3. Четкин В. А. Разработка приемника-декодера сигналов стандарта ADS-B: магистерская диссертация // Уральский федеральный университет имени первого Президента России Б. Н. Ельцина, Институт радиоэлектроники и информационных технологий-РТФ, Кафедра высокочастотных средств радиосвязи и телевидения. Екатеринбург, 2014. С. 105 с. Библиогр.: С. 82–83 (18 назв.).

РЕАЛИЗАЦИЯ ПРОТОКОЛА ПЕРЕДАЧИ ДАННЫХ В БОРТОВОЙ СЕТИ СВЕРХМАЛОГО КОСМИЧЕСКОГО АППАРАТА

**В.А. Прохоров, М.А. Павлышко, А.В. Никифоров, А.М. Огурцов,
А.А. Шандицев**

*Белорусский государственный университет, Минск, Беларусь
E-mail: prokhorov@bsu.by*

На основе решений для бортовой системы управления научно-образовательного наноспутника БГУ BSUSat-2, успешно функционирующего на околоземной орбите, разработаны библиотека и протокол интерфейса CAN для взаимодействия модулей созданной инженерной модели сверхмалого космического аппарата (СМКА), позволяющие выполнять отладку бортовых систем на этапе ввода в эксплуатацию и проводить исследовательские эксперименты в учебном процессе при подготовке специалистов аэрокосмической отрасли. Средства разработки: CMake, C, Python.

Ключевые слова: СМКА; BSUSAT-2; CAN.

Ввиду сложности передачи больших объемов данных по интерфейсу CAN и для обеспечения модульности и совместимости всех бортовых систем СМКА, в ходе разработки BSUSat-2 были созданы специальные библиотека и протокол для взаимодействия всех бортовых систем СМКА посредством CAN-шины [1,2].

Протокол взаимодействия решает сразу несколько задач: передачу больших объемов информации, унификацию взаимодействия устройств на шине, рассылку широковещательных сообщений, работу с конкретным устройством, одновременную асинхронную обработку сообщений, в том числе, от одного отправителя. Библиотека, обеспечивающая поддержку протокола, была разработана на языке C и может использоваться на любом устройстве, поддерживающем язык C.

Основная идея заключается в разбиении сообщения на части, которые передаются в шину CAN, после чего устройство, которому адресовано сообщение, получает CAN-кадры и восстанавливает из них исходное сообщение. На рис. 1 представлен заголовок протокола передачи данных.

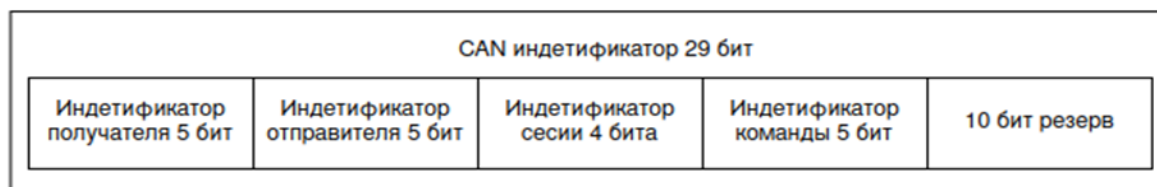


Рис. 1. Заголовок протокола передачи данных

В системе зарезервированы следующие команды для пиринговой передачи: начало передачи (START), передача данных (DATA), конец передачи (END), остановка (ABORT), подтверждение приема команды (ACK), отклонение команды (NACK).

Для работы в широковещательном режиме зарезервированы только три команды: начало передачи (NSTART), передача данных (NDATA), конец передачи (NEND). Ответы от конечных получателей в данном режиме не предусматриваются. На рис. 2 представлено схематическое изображение сессии передачи данных.

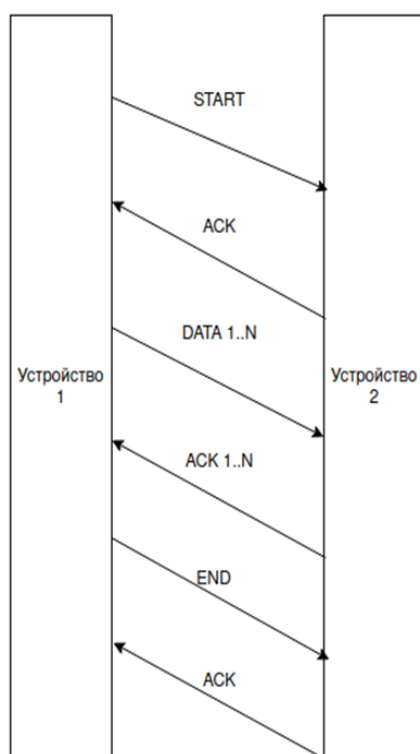


Рис. 2. Сессия передачи данных

Отправитель инициирует начало передачи, а получатель подтверждает готовность принять сообщение. В случае, если начало передачи невозможно по причине сбоя на стороне получателя, он может отклонить начало передачи (превышен лимит активных сессий и т. д.). Далее отправитель передает пакеты с данными, а получатель подтверждает успешный прием каждого пакета. Когда все данные переданы получателю, отправитель закрывает сессию. В случае сбоя получатель и отправитель могут прервать сессию передачи.

Подходы, заложенные в протокол, позволяют использовать все средства работы интерфейса CAN. Например, фильтрацию идентификаторов CAN-кадров, что позволяет игнорировать кадры, адресованные другим

устройствам на аппаратном уровне и уменьшить потребление вычислительных ресурсов. Это обеспечивается тем, что заголовок протокола располагается в идентификаторе CAN-кадра.

Также в библиотеке, реализующей данный протокол, обеспечивается возможность тонкой настройки для реализации максимальной производительности на конкретном устройстве. Пользователь библиотеки может выбрать тип выделения памяти: статический или динамический, идентификатор устройства, максимальную длину сообщения, количество буферов на отправку и прием сообщений, количество одновременно открытых сессий, время ожидания кадра в рамках сессии.

Тестирование библиотеки проводилось в виртуальной среде, структурная схема которой представлена на рис. 3.

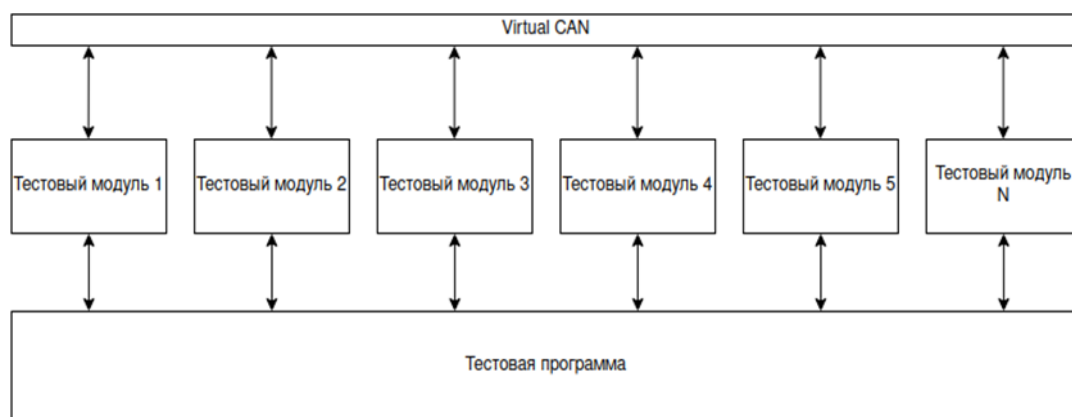


Рис. 3. Структурная схема тестовой виртуальной среды

Каждый программный модуль, подключенный к сети *Virtual CAN*, имитирует работу конечного устройства, тестовая программа следит за работой каждого модуля. Это позволяет значительно упростить отладку исходного кода и эмулировать ситуации, которые сложно воспроизвести в целевой системе, а также проверить работу библиотеки в предельных режимах работы.

ВЫВОД

Представленные в статье протокол и библиотека были протестированы на различных аппаратных платформах (STM32L4, Raspberry Pi, PC, ATSAM), успешно функционируют в BSUSat-2 и разработанной инженерной модели СМКА.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. CAN Specification Version 2.0 // Robert Bosch GmbH, 1991. P.72.
2. EMR 3 CAN BUS specification Version 11-3//DEUTZ, 2005. P.53.

ЛАБОРАТОРНЫЙ СТЕНД ДЛЯ МОНИТОРИНГА УДАЛЕННЫХ ОБЪЕКТОВ С ИСПОЛЬЗОВАНИЕ ПРОТОКОЛА AX.25

Д. В. Резниченко, И. А. Шалатонин, А. А. Спиридонов,
М. В. Гранковская

Белорусский государственный университет, Минск, Беларусь
E-mail rct.rezniche@bsu.by

Предложен вариант мониторинга удаленных объектов посредством сбора, обработки телеметрии и, при наличии необходимого функционала на принимающей стороне, отправкой корректирующих данных. Приведено описание и перечислены преимущества любительского пакетного протокола радиосвязи AX.25. Описаны принцип осуществления приема и передачи данных и устройство приемопередатчика.

Ключевые слова: микроконтроллеры; микропроцессорные системы; AX.25; IoT; радиосвязь; телеметрия; программно-определяемая радиосистема (SDR).

С развитием технологий и распространением интернета вещей (IoT), в особенности промышленного интернета вещей (IIoT), встал вопрос о создании систем, позволяющих организовать эффективный сбор и обработку данных с устройств для своевременного реагирования на инциденты и планирования дальнейшей работы производства с опорой на актуальные сведения. Решения данной задачи предполагает развитие методов удаленного мониторинга. В рамках работы рассмотрен один из методов реализации подобной системы с использованием протокола AX.25.

Протокол AX.25 работает на канальном уровне (уровень 2) модели OSI. Однако для успешной установки связи его реализация опирается на аспекты как физического уровня (уровень 1), так и, в меньшей степени, сетевого уровня (уровень 3). Рассмотрим основные параметры протокола на данных уровнях. Физический уровень в передаче данных отвечает за процесс формирования изменений уровней напряжения на передатчике для последующей модуляции. Прием сообщений представляет собой обратный процесс преобразования изменения уровней напряжения от передатчика в поток исходных данных. Протокол AX.25 не зависит от конкретной физической среды, используемой для передачи. Он может работать по различным каналам, используемым в любительской радиосвязи, включая диапазоны ОВЧ/УВЧ, спутниковые каналы и даже модуляцию звуковой карты для пакетного радио через Интернет. Физический уровень реализован согласно широко распространенному стандарту G3RUH для скорости передачи данных 9600 bps [1].

На канальном уровне протокол AX.25 использует три вида кадров для коммуникации: нумерованные (N) кадры (используются для установления/завершения соединения, отправки управляющей информации); супер-

визорные (S) кадры (используются для базового управления потоком данных и устранения ошибок в рамках связи в режиме соединения); информационные (I) кадры (основные кадры AX.25, несут фактическую полезную нагрузку данных, предназначенную для станции-получателя. Их можно использовать как в подключенном, так и в режиме без установления соединения).

На сетевом уровне протокол AX.25 не поддерживает сетевую маршрутизацию для доставки сложных пакетов через несколько переходов. Однако его можно использовать в качестве носителя для протоколов более высокого уровня, таких как NET/ROM или даже TCP/IP.

Протокол AX.25 широко используется для связи с малыми космическими аппаратами (МКА) и построения сетей пакетной передачи данных и обладает следующими достоинствами:

- каждый пакет AX.25 включает позывные отправителя и получателя, что обеспечивает автоматическую идентификацию станции во время передачи;
- AX.25 предлагает режимы как с установлением соединения, так и без него, что позволяет адаптировать его к различным потребностям связи;
- AX.25 может выступать в качестве носителя для протоколов более высокого уровня, таких как TCP/IP и NET/ROM;
- AX.25 включает поле дигипитера в структуру пакета. Дигипитеры – это станции, автоматически ретранслирующие пакеты, расширяя диапазон передачи за пределы прямой видимости между отправителем и получателем [2].

Для изучения протокола AX.25 и приобретения навыков работы с SDR трансиверами нами предложен лабораторный стенд, структурная схема которого представлена на рис. 1. Стенд прост в изготовлении и включает в себя приемопередатчик, работающий по протоколу AX.25, и комплект датчиков, поддерживающих протоколы I2C, SPI, подключаемых через отладочную плату STM32F429 disco. Стенд формирует кадры передачи телеметрии на станцию приема, выполненную на основе SDR-приемника и свободного программного обеспечения: SDR++, UZ7HO Sound Modem, AGW Online KISS [3].

Для реализации системы связи был использован отладочный комплект компании «Терраэлектроника» TE-CC430EV-433 на базе радиочастотного модуля TE-CC430F51-433 и отладочной платы MSP430-EXP430G2 (LaunchPad), производства компании «Texas Instruments (TI)». На рис. 2 приведена структурная схема отладочного комплекта. Микросхема CC430F5137 представляет собой систему-на-кристалле (СнК), которая состоит из микроконтроллерного ядра MSP430F5xx и многоканального RF-трансивера с низким энергопотреблением CC1101.

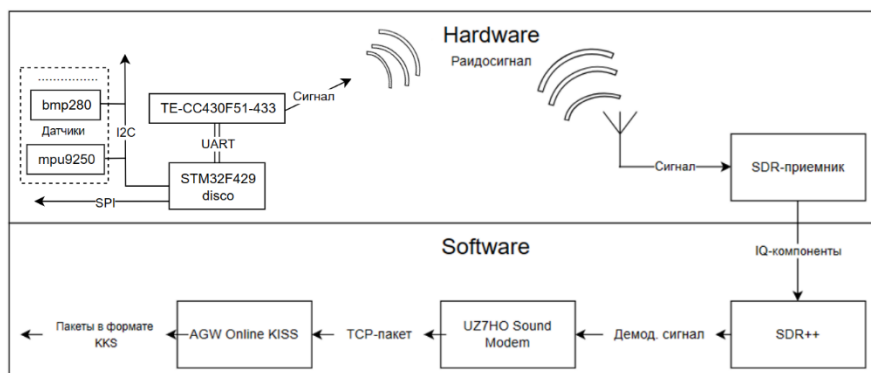


Рис. 1. Структурная схема системы приема-передачи данных

Микроконтроллерное ядро MSP430F5xx, входящее в состав СнК CC430F5137, имеет максимальную тактовую частоту 25 МГц, 12-разрядный аналого-цифровой преобразователь (АЦП) с частотой выборок 200 kSPS, набор последовательных интерфейсов (UART, 2xSPI, I2C, IrDA) и аппаратный модуль шифрования AES-128. Данный модуль предназначен для работы в безлицензионном ISM-диапазоне 433 МГц. [4]

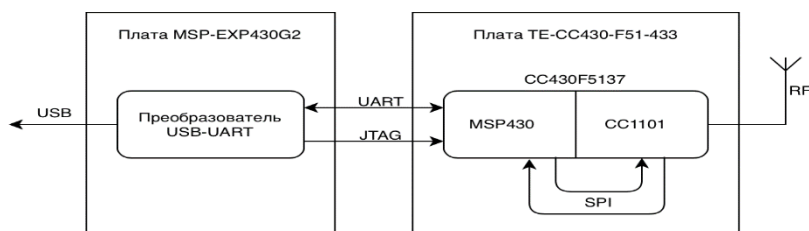


Рис. 2. Структурная схема отладочного комплекта TE-CC430EV-433

В AX.25 протоколе используется бит-ориентированный протокол канального уровня сетевой модели OSI: High-Level Data Link Control (HDLC), согласно которому кадры передаются кодированные модифицированным потенциальным кодом.

ПО микроконтроллера MSP430 должно реализовать алгоритм функционирования пакетного контроллера в KISS режиме работы:

- получать от радио приемо-передатчика CC1101 демодулированные цифровые сигналы, декодировать и дескремблировать эти сигналы, выполнять разбор HDLC пакета, проверять целостность пакета и передавать на главный компьютер полезное содержимое по KISS-протоколу;
- получать от главного компьютера предназначенную для передачи информацию по KISS-протоколу, вычислять контрольную сумму, осуществлять формирование HDLC-пакета с последующим его скремблированием и кодированием, передавать полученный пакет на радио приемопередатчик CC1101.

На рис. 3 представлена структура разработанного ПО, удовлетворяющего перечисленным требованиям. Пунктирными стрелками отмечен путь, который проходят данные, передаваемые с главного компьютера, до того момента как попадут в эфир. Сплошными стрелками отмечен обратный путь.

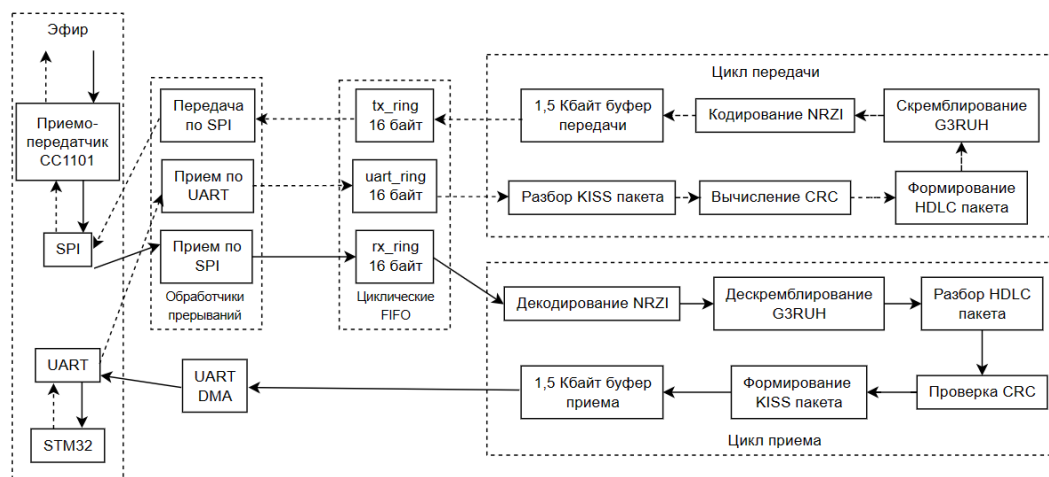


Рис. 3. Структурная схема ПО пакетного контроллера в KISS

Предложенный учебный комплекс позволяет изучить протокол AX.25, подключать новые датчики и приобрести навыки работы с SDR трансиверами. Учебный лабораторный комплекс используется при подготовке специалистов на кафедрах физики и аэрокосмических технологий и телекоммуникаций и информационных технологий.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Бучинский Д.А., Семенович С. Н. Реализация физического уровня протокола любительской пакетной радиосвязи AX.25 университетского микроспутника [Электронный ресурс]. URL: <https://elib.bsu.by/handle/123456789/159737> (дата обращения: 30.03.2024).
2. Beech A., Nielsen E., Taylor J. AX.25 Link Access Protocol for Amateur Packet Radio // 1998. P. 133.
3. Учебно- методический комплекс на базе программно-аппаратных средств для лабораторной отработки нано и пикоспутников / Шалатонин И.А., Спиридонов А.А., Саечников В.А. [и др.] // Комплексные проблемы техносферной безопасности: материалы Международной научно - практической конференции Воронеж: ФГБОУ ВПО «Воронежский государственный технический университет», 2017. Ч. IV. 250 с. С. 54-58
4. Радиочастотный модуль TE-CC430F51-433 [Электронный ресурс]. URL: <https://www.elmatrix.ru/upload/docs/te-cc430f51-433.pdf?ysclid> (дата обращения: 26.05.2024).

РАЗРАБОТКА ПРОГРАММНОГО КОМПЛЕКСА МОНИТОРИНГА И ОБРАБОТКИ ДАННЫХ ПО СТАНДАРТУ ISO 11898 ДЛЯ ЭЛЕКТРИЧЕСКИХ ТРАНСПОРТНЫХ СРЕДСТВ

С. Д. Рундыгин, А. А. Субач, А. Л. Труханович

Белорусский государственный университет, Минск, Беларусь

E-mail: rct.rundygin@bsu.by

Доклад описывает разработку программного комплекса для мониторинга и обработки данных в электрических транспортных средствах с использованием стандарта ISO 11898. В работе представлены выбранные программно-аппаратные средства, основные модули приложения и их функциональность. Рассматривается пример подключения каналов управления и отображение полученных данных для электродвигателя и аккумуляторной батареи. Результатом разработки стало приложение, обеспечивающее получение, обработку и отображение данных в режиме реального времени.

Ключевые слова: электрические транспортные средства; ISO 11898; программный комплекс; мониторинг; обработка данных; визуализация.

Развитие электрических транспортных средств (ЭТС) является одним из ключевых направлений в автомобильной индустрии. Электрические автомобили и другие ЭТС предлагают экологически более чистые и энергоэффективные альтернативы автомобилям с двигателями внутреннего сгорания.

Диагностика стала неотъемлемой частью современного автомобиля, так как электронные системы, встроенные в автомобиль, становятся все сложнее. Они включают в себя множество микроконтроллеров, которые взаимодействуют друг с другом для обеспечения различных функций. Поэтому важно иметь систему, способную собирать и анализировать информацию о состоянии автомобиля. Так как она позволит обнаруживать и сообщать о возможных неисправностях, используя стандартизованные коды ошибок. Благодаря стандартизации, собранные данные могут быть анализированы с помощью специальных устройств [1].

CAN-интерфейс является стандартом коммуникации между различными компонентами в автомобиле и обеспечивает передачу информации о состоянии различных систем, таких как двигатель, тормозная система, аккумуляторная батарея и другие. ISO 11898 – это международный стандарт, который определяет физический и логический уровни для передачи данных по CAN-шине в автомобилях и промышленных приложениях [1, 2]. Поэтому стоит вопрос о создании программного обеспечения, которое можно интегрировать в ЭТС для считывания и обработки данных используя CAN-интерфейс.

Отслеживать состояния ЭТС можно двумя способами: встроенная диагностика и с использованием диагностического сканера. Встроенная диагностика отображает данные, которые были заложены производителем. А способ с использованием диагностического сканера нуждается в дополнительном устройстве(ноутбук/телефон), на которое будет установлено специальное ПО для обработки и визуализации данных [3, 4].

В следствии этого возникает потребность в создании комплекса, который будет легко масштабироваться (иметь возможность добавления обработки требуемых данных) и содержать в себе необходимые средства для обработки данных (ПО) и их визуализации.

Для разработки приложения выбраны следующие программно-аппаратные средства: Raspberry PI 4 8GB [5] с использованием операционной системы Raspberry OS (на базу Linux Debian), для работы с CAN-шиной применяется модуль RS485 CAN HAT [6] (на основе CAN контроллера MCP2515 [7]).

Приложение состоит из трех основных модулей:

1. Модуль CAN-шины выполняет следующие функции: отправка запросов данных в CAN-шину, проверка буфера сообщений, чтение данных из буфера сообщений и отправка их в модуль обработки данных.

Алгоритм работы модуля CAN-шины состоит в получении данных от устройств через CAN-интерфейс (на основе ISO 11898). Эти данные могут быть запрошены или автоматически получены с заданным периодом для каждого сообщения. Информация извлекается из буфера сообщений и передает ее в модуль обработки данных [8].

2. Модуль обработки данных имеет следующий функционал: прием данных от модуля CAN-шины, обработка данных в соответствии с их типом, отправка данных в модуль графического интерфейса.

Модуль обработки данных при получении сообщения от модуля CAN-шины, используя сдвиг и размер полезных данных соответствующий конкретному сообщению, получает нужную информацию, обрабатывает её, в зависимости от типа сообщения, и передает в модуль графического интерфейса. Для нахождения полезных данных используется отдельный масштабируемый класс. Этот механизм позволяет легко расширять список CAN сообщений и тем самым добавлять новые устройства в систему.

3. Модуль ГИ занимается приемом информации от модуля обработки данных и отображением заданных параметров.

Основной цикл работы приложения представлен на рис. 1. При подключении к блоку управления (БУ) электродвигателя (ЭД) и батареи через RS485 CAN HAT структура системы будет выглядеть следующим образом (рис. 2).

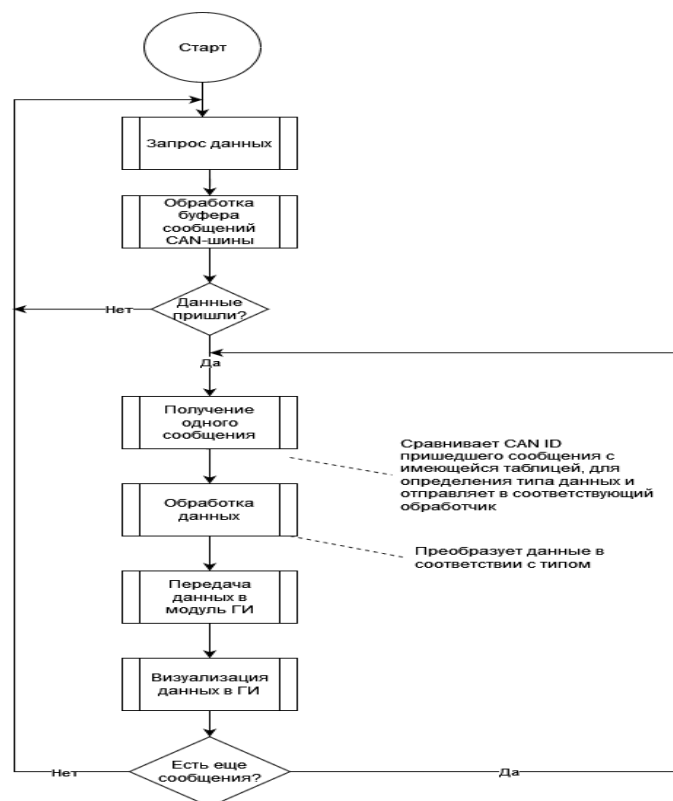


Рис. 1. Блок-схема работы основного цикла приложения.

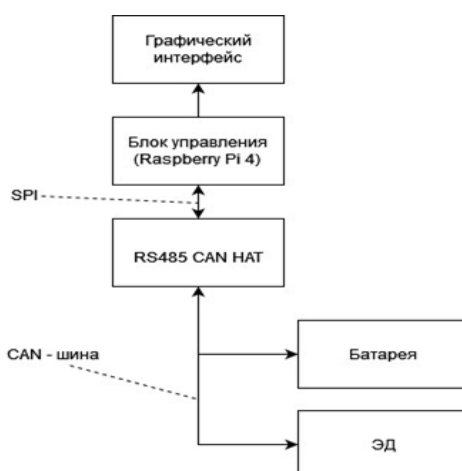


Рис. 2. Схема подключения компонентов к CAN-шине.

С ЭД получаем скорость, обороты, пробег, режим движения и ошибки. С батареи получаем заряд, состояние зарядки, температуру ячеек. Заряд батареи позволяет рассчитать примерное расстояние, которое ЭТС сможет проехать до полной разрядки. Температура батареи и ЭД сигнализирует о перегреве. Ошибки ЭД сигнализируют о неисправностях и их характере.

На рис. 3 представлен скриншот приложения с отображаемыми данными.

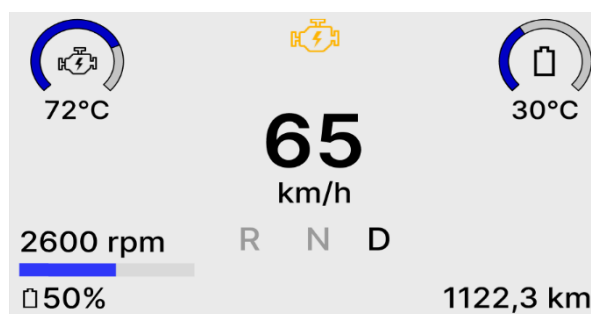


Рис. 3. Скриншот приложения.

В результате разработана структура системы мониторинга и обработки данных по стандарту ISO 11898 для ЭТС, на основе которой было создано приложение, позволяющее в режиме реального времени получать данные по CAN-интерфейсу, обрабатывать их и визуализировать. Это позволяет пользователю контролировать состояние машины и вовремя замечать неисправности, которые могут привести к аварии.

Преимуществом системы является ее гибкость и расширяемость: можно настраивать под различные компоненты и добавлять новые устройства. Для этого достаточно добавить поля в класс с CAN сообщениями и создать новый элемент ГИ в приложении для отображения информации.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Nicolas Navet, Françoise Simonot-Lion. Automotive Embedded Systems Handbook.
2. ISO 11898 [Электронный ресурс]. URL: <https://files.stroyinf.ru/Data/610/61075.pdf> (дата обращения: 14.03.2024).
3. Автомобильная электроника [Электронный ресурс]. URL: <https://books.google.by/books?id=dgaEw-MYnn9oC&printsec=frontcover&hl=ru#v=onepage&q&f=false> (дата обращения: 18.04.2024).
4. Автотроника [Электронный ресурс]. URL: https://fileskachat.com/view/50245_8cf2c5d0cf3d02ca3ac890d209801622.html (дата обращения: 15.04.2024).
5. Raspberry Pi 4 Computer Model B Manual [Электронный ресурс]. URL: <https://datasheets.raspberrypi.com/rpi4/raspberry-pi-4-datasheet.pdf> (дата обращения: 14.03.2024).
6. RS485 CAN HAT Manual [Электронный ресурс]. URL: <https://www.waveshare.com/w/upload/2/29/RS485-CAN-HAT-user-manuakl-en.pdf> (дата обращения: 14.03.2024).
7. MCP2515 Datasheet [Электронный ресурс]. URL: <https://ww1.microchip.com/downloads/en/DeviceDoc/MCP2515-Stand-Alone-CAN-Controller-with-SPI-20001801J.pdf> (дата обращения: 14.03.2024).
8. CAN communication on the Raspberry PI with SocketCAN [Электронный ресурс]. URL: <https://www.pragmaticlinux.com/2021/10/can-communication-on-the-raspberry-pi-with-socketcan/> (дата обращения: 14.03.2024).

АЛГОРИТМ ДЛЯ ОТЛАДКИ ПРОЦЕССОВ ВЗАИМНОЙ СИНХРОНИЗАЦИИ УЗЛОВ MESH-СЕТИ

А.С. Семенович, С.Н. Семенович, И.П. Стецко

Белорусский государственный университет, Минск, Беларусь

E-mail: semenoviAS1@bsu.by

В работе исследуется алгоритм взаимной синхронизации устройств mesh-сети, основанный на рекламных пакетах (Advertising PDU) в сочетании с функционалом исходного стека BLE.

Ключевые слова: Bluetooth Low Energy; время синхронизации; беспроводные сети; Интернета вещей; IEEE 802.1AS.

Синхронизация времени является ключевой функцией для беспроводных сетей из сенсорных устройств, которые применяются в медицинских и спортивных IoT-приложениях, часто подразумевающих использование большого количества автономных устройств, объединённых в одну беспроводную mesh-сеть, узлы которой взаимодействуют друг с другом, обмениваясь данными, и выполняют различные асинхронные задачи [1]. При использовании ячеистых сетей со сложной маршрутизацией данных требуется взаимная синхронизация внутренних таймеров всех устройств для обеспечения хронологического порядка при регистрации событий от встроенных датчиков, для координации этапов общего измерительного процесса между устройствами и синхронизации исполнительных устройств сети.

Хотя в сенсорных сетях с ячеистой топологией используются различные беспроводные протоколы, именно Bluetooth Low Energy в последние годы привлек внимание разработчиков благодаря аппаратной поддержке режимов коммуникации с низким энергопотреблением и доступности готовых решений для автономных датчиков.

В данной работе исследован алгоритм взаимной синхронизации устройств mesh-сети, основанный на рекламных пакетах (ADV PDU) в сочетании с функционалом исходного стека BLE. Продемонстрирован метод оценки смещения хронологии событий и дрейфа частоты встроенного таймера узлового устройства, а также исследовано влияние количества пакетов синхронизации на общую точность взаимной синхронизации узлов.

Предложенный алгоритм синхронизации основан на использовании уровня MD (media-dependent) стандарта IEEE 802.1AS для беспроводных систем связи. Стандарт определяет протокол и процедуры, используемые для обеспечения синхронизации времени с минимальным дрейфом синхронизации в течении суток для мостовых и виртуальных мостовых локальных сетей с критически важной хронологией событий. MD-сущность

отвечает за передачу временных меток и синхронизацию времени между устройствами по принципу «master-slave» – подчиненные узлы синхронизируются сами с часами главного узла на разных уровнях mesh-сети. Такой подход обеспечивает субмикросекундную точность синхронизации без потери пропускной способности радиоканала для приложений с асинхронными процессами [1].

Алгоритм взаимной синхронизации WPTP (Wireless precision time protocol) выполняет синхронизацию по уровням [2]. На первом этапе синхронизируются узлы одного уровня. Эти узлы становятся главными для узлов следующего уровня только после синхронизации узлов предыдущего уровня, и так далее, как показано на рис. 1.

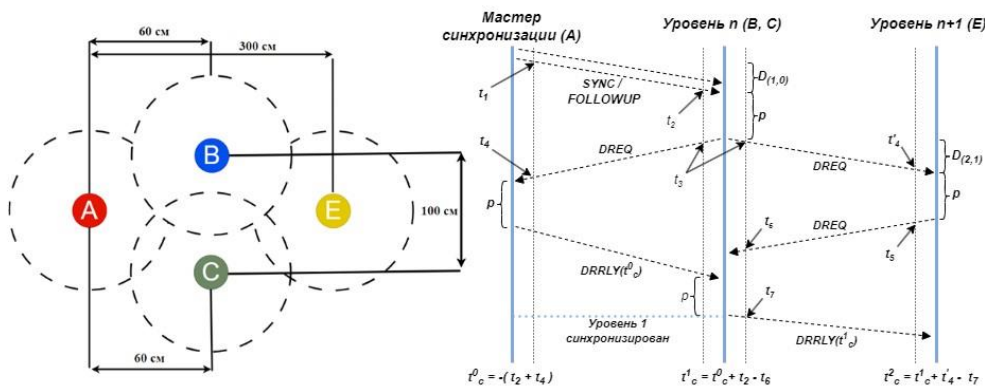


Рис. 1. Принцип работы алгоритма взаимной синхронизации узлов mesh-сети

Таким образом, ведомые узлы могут быть соединены с ведущим двумя способами: *напрямую* – узлы, которые находятся в пределах диапазона передачи главного узла т.е. узлы уровня n (узлы В и С); *косвенно* – узлы, которые не находятся в пределах диапазона передачи главного узла (узел Е). Для этого используются промежуточные узлы в качестве переходов для соединения с ведущим (т.е. узлы уровня n).

Синхронизация узлов n уровня. Формула для вычисления смещения выглядит следующим образом:

$$\Delta = \frac{(t_2 + t_3) - (t_4 + t_1)}{2} \quad (1)$$

где t_2, t_3 относятся к шкале времени ведомого устройства, временные метки t_1, t_4 относятся к шкале времени мастер-устройства, следовательно они могут быть рассчитаны на этом уровне и отправлены в виде единой консолидированной временной метки t_c^0 для последующего использования ее на уровне n , где время смещения события будет рассчитываться по формуле (2):

$$\Delta_{(1,0)} = \frac{(t_2 + t_3) + t_c^0}{2} \quad (2)$$

Следовательно, алгоритм синхронизации требует только три радиопакета для вычисления и коррекции смещения события. Для учета задержки передачи сообщения используется второй набор сообщений: *SYNC*-маяк используется для инициации синхронизации и *FOLLOWUP*-маяк, используемый для поддержания синхронизации.

Синхронизация узлов n+1 уровня. Взаимная синхронизация узлов осуществляется по Bluetooth в ISM-диапазоне 2,4 ГГц (полоса частот которого разделена на 40 физических каналов, только три из которых (№ 37, 38 и 39) отведены под рекламные сообщения), путем обмена сообщениями синхронизации (маяками) в несколько этапов [3].

Использования ADV-каналов связи в беспроводных сетях подразумевает, что каждое сообщение-маяк будет передаваться как широкополосное. Следовательно, *DREQ*-маяк запрос данных у рекламирующего устройства, отправленное узлами уровня n узлам предыдущего уровня n-1 может быть использовано как *SYNC*-маяк для узлов следующего уровня n+1. Тогда смещение событий для узлов уровня n+1 вычисляется по формуле:

$$\Delta_{(2,0)} = t'_4 - D_{(2,1)} - p - D_{(1,0)} - t_1 \quad (3)$$

где $D_{(2,1)}$ – задержка передачи между уровнем n+1 и уровнем n; p – время на прием и обработку сообщения узлом, предполагается одинаковым для всех узлов; $\Delta_{(2,0)}$ – смещение часов между узлами уровня n+1 и уровня 0. Используя терминологию временных меток (рис. 1) и формулу (3) для расчета задержки передачи сообщения между уровнями, можно вычислить временную метку, вносимую на каждом уровне:

$$\Delta_{(2,0)} = \underbrace{(t'_4 + t_5)}_{\text{уровень } n+1} + \underbrace{(t_2 - t_6)}_{\text{уровень } n} + \underbrace{(t_2 - t_6)}_{\text{уровень } 0} + \overbrace{t_c^1} \quad (4)$$

Так как уровень 0 уже был рассчитан, следовательно можно передавать единую консолидированную временную метку $t_c^1 = (t_2 - t_6) + t_c^0$ до уровня n+1. Результирующая формула для вычисления смещения на уровне n+1 будет выглядеть как:

$$\Delta_{(n,0)} = \frac{1}{2}(t_{r(n-1)}^n + t_{s(n-1)}^n + t_c^{n-1}) \quad (5)$$

где t_c^{n-1} может быть рассчитан как:

$$t_c^{n-1} = \begin{cases} t_c^{n-1} + t_r^n - t_{r(n+1)}^n, & \text{если } n > 1 \\ -(t_4 + t_1), & n = 0 \end{cases} \quad (6)$$

Таким образом, алгоритм взаимной синхронизации узлов mesh-сети построен на передачи консолидированного значения временной метки узлам уровня $n-1$, используя *DRPLY*-маяк, в то время как маяки *SYNC* и *DREQ* не содержат временных меток и используются только для инициирования взаимной синхронизации. Такой механизм значительно сокращает общее время согласования временных шкал и количество пакетов синхронизации без ущерба для точности синхронизации и пропускной способности канала связи.

На рис. 2 (слева) представлены осциллограммы взаимной синхронизации трех устройств разных уровней в режиме свободно работающего таймера – сигналы синхронизации не передаются; синим цветом показан мастер синхронизации; красным – узлы 1 уровня; зеленым – узлы 2 уровня. В течение 1 часа максимальный разброс значений для смещения события 3 уровня составил ~ 700 нс. Справа на рис. 2 – результат эксперимента с использованием сигналов синхронизации, в этом случае наибольший разброс составил ~ 100 нс.

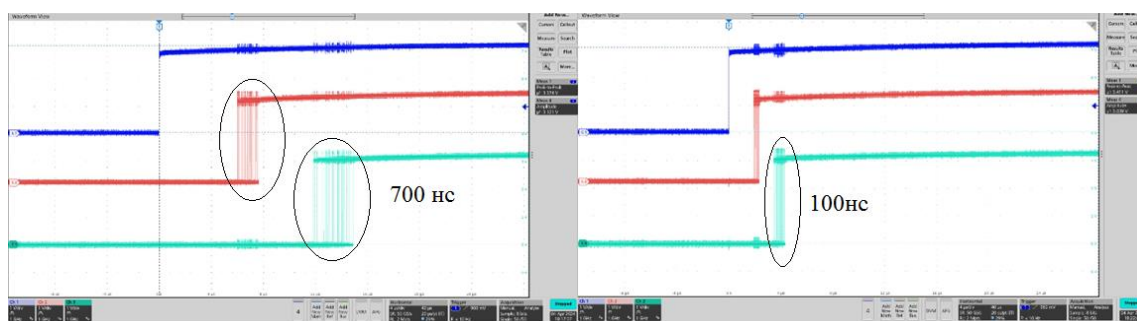


Рис. 2. Осциллограммы смещений событий синхронизации для узлов 3-х уровней

ВЫВОД

Была исследована задача отладки процессов взаимной синхронизации узлов BLE mesh-сети на основе протокола WPTP. Проведен эксперимент, подтверждающий высокую точность взаимной синхронизации и возможность применения этого алгоритма в IoT-приложениях для реализации измерительного процесса и синхронизации исполнительных устройств сети в режиме реального времени.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Семенович А.С. Масштабируемый IoT тренажер спортсмена. Сборник научных работ студентов Республики Беларусь «НИРС 2020». 2021. С. 143-144.
2. IEEE 1588 for Clock Synchronization in Industrial IoT and Related Application: A Review on Contributing Technologies, Protocols and Enhancement Methodologies / Idrees, Z., Granados, J., Sun, Y. [et al.] // IEEE Access. 2020.
3. Iannizzotto, G.; Milici, M.; Nucita, A.; Lo Bello, L. A Perspective on Passive Human Sensing with Bluetooth. Sensors. 2022.

ФУНКЦИОНАЛЬНАЯ СТРУКТУРА ИНЖЕНЕРНОЙ МОДЕЛИ СВЕРХМАЛОГО КОСМИЧЕСКОГО АППАРАТА

С.Н. Семенович, И.П. Стецко, А.М. Огурцов, Д.Г. Терешко, А.В. Калачёва, Т.А. Жилинская, Е.А. Барсуков

*Белорусский государственный университет, Минск, Беларусь
E-mail: semenovich@bsu.by*

Разработана инженерная модель сверхмалого космического аппарата, компоновка и функциональная структура которого позволяет оперативно менять состав используемых модулей, обеспечивать их взаимодействие и управление работой в процессе проведения учебных и исследовательских экспериментов. Средства разработки: Micropyton.

Ключевые слова: СМКА; микроконтроллер; CAN; I²C.

Инженерная модель сверхмалого космического аппарата (ИМ СМКА), разработанная для образовательных применений, выполнена в форм-факторе Cubesat 3U [1], удобном при использовании в учебных лабораториях. Внутри продольной грани каркаса ИМ СМКА закреплена объединительная электронная плата бортовой системы энергоснабжения и коммуникации, в опорной плоскости которой размещены 10 жёстких разъёмных соединений для подключения электронных печатных плат функциональных модулей. Такая «слотовая» компоновка предоставляет возможности комбинирования набором функциональных модулей для работы различных учебных и исследовательских задач.

В функциональной структуре ИМ СМКА, представленной на рис. 1, можно выделить следующие функциональные элементы:

- «материнскую» бортовую систему энергоснабжения и коммуникации, включая блоки контроля мощности, мониторинга и распределения энергии, бортовые коммуникационные шины, а также модули солнечных батарей и функциональные модули аккумуляторных батарей;
- функциональный модуль системы бортового компьютера со встроенным электронным модулем регистрации изображения и измерительных датчиков;
- функциональные модули бортовой системы радиосвязи;
- систему определения ориентации и стабилизации ИМ СМКА, включая функциональный модуль электромагнитной ориентации и стабилизации, модули датчиков направления на Солнце и системы оптического трекинга;
- функциональный модуль системы радионавигации ИМ СМКА со встроенным электронным модулем датчиков расстояния и лазерного указателя.



Рис. 1. Структурная схема инженерной модели сверхмалого космического аппарата

Бортовая система энергоснабжения и коммуникации (БСЭиК) ИМ СМКА обеспечивает электропитание других систем и модулей ИМ СМКА, а также коммуникацию между ними. Структурная схема внутренних взаимосвязей БСЭиК, с точки зрения энергетических потоков и узлов управления, представлена на рис. 2.

Структуру БСЭиК составляют следующие функциональные узлы:

- блок контроля мощности (БКМ) с модулями солнечной и аккумуляторных батарей;
- блок мониторинга и распределения энергии (БМРЭ) с коммутаторами энергопитания (КЭ);
- бортовые коммуникационные шины.

БСЭиК, в частности, интегрирует внутреннюю интерфейсную шину СМКА для выдачи всей необходимой телеметрии в бортовую систему управления спутника.

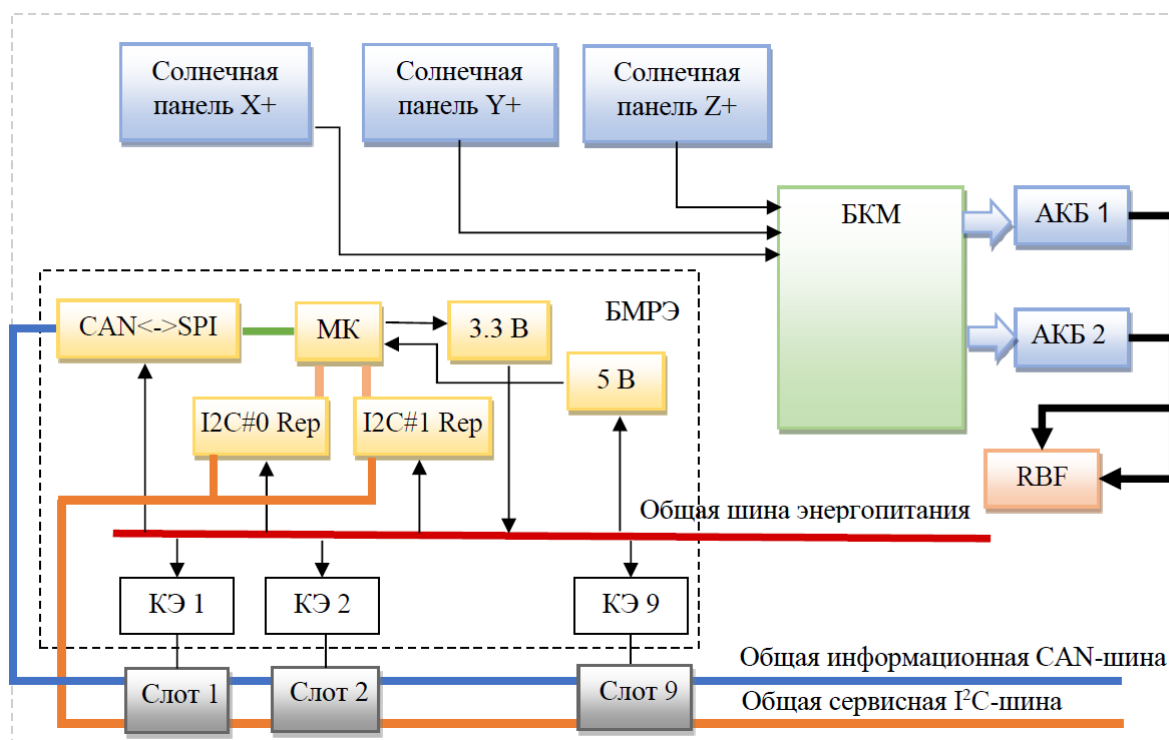


Рис. 2. Структурная схема внутренних взаимосвязей системы энергоснабжения и коммуникации СМКА

Блок контроля мощности (БKM) обеспечивает поддержание штатного режима работы системы энергообеспечения СМКА при преобразовании солнечной энергии в электрическую и её накоплении в основной и резервной бортовых аккумуляторных батареях. Основу БKM составляет программно-управляемая платформа многоканального контроля силы постоянного тока, имеющая цифровой интерфейс I^2C .

Основу блока мониторинга и распределения энергии (БМРЭ) составляет микроконтроллер (МК) *RP2040* на платформе разработки *Raspberry Pi Pico*. В составе БМРЭ для каждого из слотов электронных модулей ИМ СМКА имеются десять коммутаторов энергопитания (КЭ), реализованных на базе микросхемы переключателя нагрузки с ограничением по току. Контроль тока в цепи каждого слота выполняется с помощью монитора тока и напряжения с цифровым интерфейсом I^2C . В случае превышения установленного лимита энергопотребления, коммутатор закрывается и выдает адресное сообщение на шину данных, где фиксируются все события о превышении лимита. Управление КЭ и сбор информации о превышении энергетических затрат выполняется на базе двух 8-битных расширителей цифровых входов/выходов для шины I^2C , программно-управляемых от МК по шине I^2C .

В ИМ СМКА имеется общая информационная CAN–шина, программно управляемая средствами МК, являющаяся удобным средством для обмена командами и данными между всеми модулями СМКА, имеющая высокую степень надежности. Каждый модуль ИМ СМКА имеет преобразователь двунаправленной информационной шины SPI его МК в общую CAN–шину.

В дополнение к общей CAN–шине, ИМ СМКА имеет также две общие для бортовых модулей коммуникационные I²C–шины, разделенные физически. Одна I²C–шина используется для коммутации слотов и контроля параметров энергопотребления, другая выполняет функции резервной информационно-сервисной шины. Схема ретрансляции повторяется на всех электронных платах функциональных модулей ИМ СМКА.

Программное обеспечение МК разработано на языке *Micropython*. После старта оно запускает алгоритм мониторинга бортового напряжения и токов потребления, осуществляет заданную коммутацию энергетических потоков для бортовых модулей и подсистем, согласно логике алгоритма функционирования СМКА.

Реализованная система единых бортовых информационных шин CAN и I²C позволяет организовать удобное и надежное взаимодействие со всеми остальными функциональными модулями ИМ СМКА, управляемыми унифицированными МК: модулем бортовой системы радиосвязи, модулем системы бортового компьютера (основного и резервного), модулем регистрации изображения и измерительных датчиков, системой определения ориентации и стабилизации, модулем радионавигации.

ВЫВОД

Разработаны слотовая компоновка, функциональная структура бортовых функциональных модулей с унифицированными микроконтроллерами и едиными информационными шинами ИМ СМКА, позволяющие оперативно менять состав используемых модулей, обеспечивать их взаимодействие между собой и распределёнными датчиками, надёжное управление работой в процессе проведения учебных и исследовательских экспериментов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. CubeSat Design Specification [Электронный ресурс]. URL: <https://www.cubesat.org/cubesatinfo> (дата обращения: 03.04.2024).

АЛГОРИТМ ДЛЯ ОГИБАНИЯ ПРЕПЯТСТВИЙ ПРИ ДВИЖЕНИИ МОБИЛЬНОГО РОБОТА

Сидоренко А.В., Солодухо Н.А.

Белорусский государственный университет, Минск, Беларусь

E-mail: sidorenkoa@yandex.by

В работе рассмотрен модифицированный нами алгоритм перемещения робота, который позволяет ускорить его движение к цели. Алгоритм реализован на операционной системе Linux Ubuntu 22.04 с использованием ROS Humble и пакета визуализации Gazebo 11 на языках bash, Python, что позволяет роботу огибать более сложные препятствия с большей скоростью без существенных затрат вычислительной мощности.

Ключевые слова: робот; перемещение; огибание препятствий; ROS; Gazebo; Python; bash.

Введение. При изучении робототехнических задач актуальной является проблема огибания препятствий роботом при его движении. В работе [1] нами предложен алгоритм огибания препятствий мобильным роботом. Однако данный алгоритм, несмотря на отсутствие машинного обучения, что замедляло бы его работу и увеличивало требуемую вычислительную мощность, огибает препятствия не по кратчайшему пути. В то же время алгоритм VFH*, предложенный в литературе [2], обладая достаточно оптимальной траекторией при огибании препятствий, имеет следующий недостаток: при наличии достаточно длинного препятствия, расположенного перпендикулярно пути к цели, робот не способен его обойти, циклично двигаясь вправо и влево вдоль препятствия. Нами предложен алгоритм, использующий преимущества алгоритма VFH* в плане оптимальности пути огибания препятствия, и нивелирующий недостаток этого алгоритма - при наличии длинного препятствия, расположенного перпендикулярно цели движения робота, мы переключаемся на использование алгоритма, предложенного нами в работе [1].

Алгоритм VFH* и его предшественники начали широко использоваться в настоящее время для огибания препятствий при движении беспилотных летательных аппаратов, как в двумерном, так и трехмерном пространстве, что говорит об актуальности проблемы, касающейся огибания длинных препятствий, расположенных перпендикулярно пути к цели движения робота, и востребованности модификации алгоритма, которая бы решала эту проблему.

Модификация алгоритма. Суть модификации использования алгоритма состоит в следующем. Когда при огибании препятствия робот, двигаясь вдоль препятствия, начинает удаляться от цели, это заставляет его

сменить направление и двигаться вдоль препятствия в другом направлении. И если и в другом направлении, двигаясь вдоль препятствия, наступает момент, когда робот начинает удаляться от цели, он вновь разворачивается. Таким образом робот циклически двигается вдоль препятствия, не огибая его (рис. 1). Для решения данной проблемы мы модифицировали алгоритм следующим образом. Когда робот находится близко к препятствию, и начинает двигаться вдоль него, возможно, настанет момент, когда робот начнет удаляться от цели. Именно в момент, когда будет зафиксировано удаление от цели, мы модифицируем исходный алгоритм. Модификация заключается в том, что когда будет зафиксировано удаление от цели при движении вдоль препятствия, мы заставляем робота двигаться уже выбранной стороной вдоль препятствия до тех пор, пока робот не начнет приближаться к цели, огибая препятствие, двигаясь вплотную к нему. Как только робот встречает препятствие (т.е. впереди робота на расстоянии не больше 0,4 единицы координатной сетки), он поворачивается тем боком к препятствию, которым изначально был ближе к препятствию (сохраняя сторону, которой он к нему повернулся, путь это будет левый бок) и начинает двигаться вдоль препятствия по прямой до тех пор, пока изгибание препятствия не приведет к отдалению робота от препятствия на расстояние от 0,37 единицы координатной сетки. Далее робот поворачивается на 90 градусов в сторону препятствия (если робот двигался, оставляя препятствие с левого бока, он поворачивается против часовой стрелки, иначе - по часовой) и начинает движение прямо до тех пор, пока либо не отдалиться на расстояние от 0,37 единицы координатной сетки либо не встретит препятствие (т.е. приблизиться не менее чем на 0,4 единицы от координатной сетки). Если он отдалился - он снова поворачивается на 90 градусов в том же направлении что и раньше, если он приблизился - поворачивается боком к препятствию (только тем, что и раньше, т.е. в нашем случае - левым боком). Как только будет зафиксировано приближение к цели он переключается на алгоритм VFH*. В основе алгоритма VFH* лежит определение сектора, в направлении которого нет препятствий, и который был бы ближе всего к направлению к цели. На основании этого алгоритма и данных, полученных от датчиков робота, выбирается направление движения робота, такое, чтобы робот двигался среди препятствий, не сталкиваясь с ними, в то же время приближаясь к цели.

Реализация алгоритма VFH* состоит в следующем. Когда робот движется, выбирается момент, когда робот находится близко к препятствию (скажем, на расстоянии 0,37 единицы координатной сетки). Затем запускается цикл оценки секторов, доступных для движения робота на расстоянии, равном расстоянию до препятствия (в нашем случае, это 0,37 единицы координатной сетки), пошагово (с каждой итерацией увеличивая

дальность оценки доступных для движения секторов) оценивающий доступные сектора на все большем расстоянии (шаг увеличения расстояния 0,2 единицы координатной сетки).

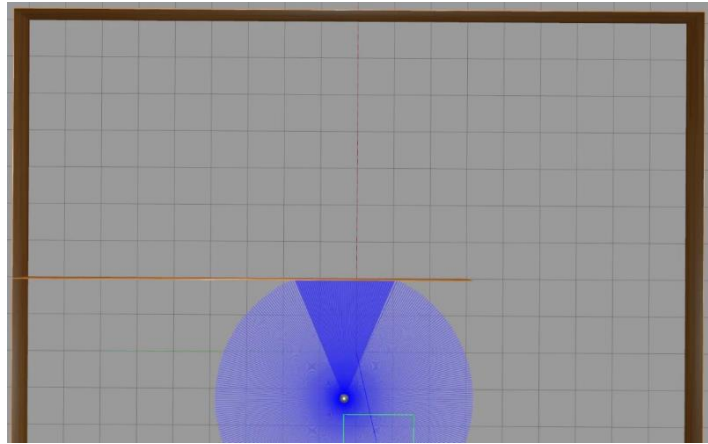


Рис. 1. Пример длинного препятствия в среде визуализации Gazebo 11

Если в процессе работы вышеуказанного цикла настал момент, когда доступных секторов нет (чаще всего это случается, когда есть закрытое пространство с поворотом), выбирается предыдущая итерация, где движение начинается в сторону наиболее удаленного препятствия. Если же в процессе работы цикла мы достигли расстояния, равного или превышающего дальность обнаружения препятствия роботом (данное расстояние равно 3,6 единицы координатной сетки), выбирается доступное для движения направление, наиболее близкое к цели. Полная структурная схема полученного алгоритма представлена на рис. 2.

При работе с симуляцией робота использован визуализатор Gazebo 11 и ROS 2 Humble Hawksbill. Для управления роботом использованы команды командной оболочки UNIX bash в ОС Linux Ubuntu 22.04 Jammy Jellyfish. Исходный код представлен в [3], ветка “vfh-algorithm”. Для работы алгоритма необходимо запустить скрипт bash lines.sh. Для задания конечной точки движения необходимо задать координаты X и Y в строках 332 и 333 файла lines.sh.

Заключение. Нами предложена модификация алгоритма, разработанного нами в работе [1]. Модифицированный алгоритм обладает более оптимальным путем движения робота, характерным для алгоритма VFH* [2], и в то же время хорошо подходит для огибания длинных препятствий, расположенных перпендикулярно движению робота по направлению к цели. Алгоритм реализован на языках bash и Python в ОС Linux Ubuntu 22.04 Jammy Jellyfish. Апробация алгоритма будет производиться в визуализаторе Gazebo 11 с использованием ROS 2 Humble Hawksbill.

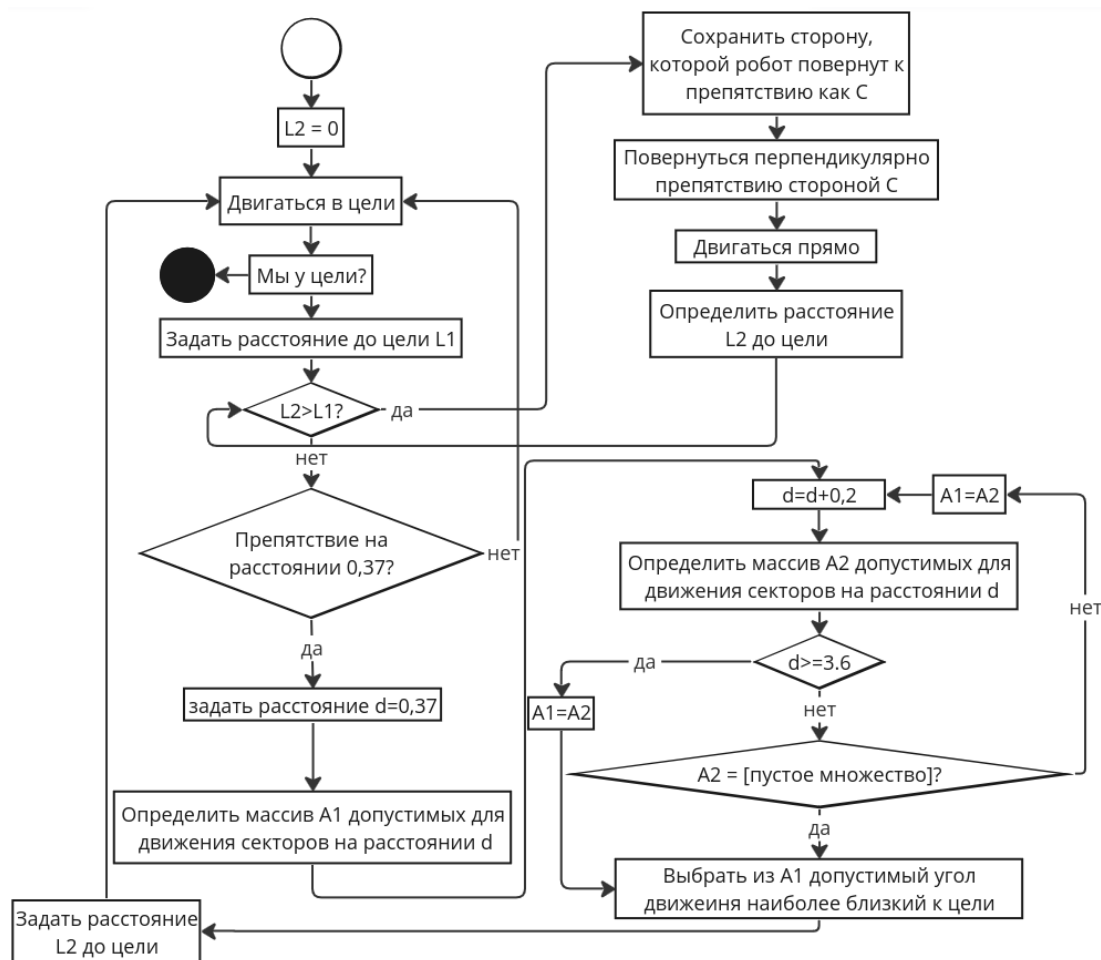


Рис. 2. Структурная схема предложенного нами алгоритма

Предполагается, что предложенный алгоритм будет обладать лучшими характеристиками по сравнению с алгоритмами в работах [1] и [2], аккумулируя преимущества этих алгоритмов и нивелируя их недостатки.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. А. В. Сидоренко, Н. А. Солодухо. Огибание препятствий при перемещении мобильного робота // Системный анализ и прикладная информатика. 2023. № 1. С. 4-9. DOI: [10.21122/2309-4923-2023-1-4-9](https://doi.org/10.21122/2309-4923-2023-1-4-9)
2. I.Ulrich, J. Borenstein. Local Obstacle Avoidance with Look-Ahead Verification // Proceedings of the 2000 IEEE International Conference on Robotics & Automation, San Francisco, CA, April 2000. San Francisco, 2000. P. 2505-2511. DOI: 10.1109/ROBOT.2000.846405
3. GitHub [Электронный ресурс]. URL: <https://github.com/MikitaSaladukha/ROSproject> (дата обращения: 19.03.2024).

ИНФРАСТРУКТУРА УМНОГО КАМПУСА

А.М. Соболев, В.П. Кочин

Белорусский государственный университет, Минск, Беларусь

E-mail: sobolam@bsu.by

В данной статье рассматривается инфраструктура умного кампуса, описаны его компоненты, связь с технологией IoT. Предложена концепция многоуровневой модели умного кампуса для решения задачи обеспечения информационной безопасности умного кампуса.

Ключевые слова: умный кампус; интернет вещей; компоненты умного кампуса; взаимодействие систем.

ВВЕДЕНИЕ

В современном мире технологии проникают во все сферы жизни, и образование не является исключением. Умный кампус – это новое поколение структуры образовательных учреждений, которые используют передовые технологии для создания инновационной и стимулирующей обучающей среды. Суть умного кампуса заключается в том, чтобы создать интеллектуальную среду, которая поддерживает активное обучение, стимулирует творческое мышление и обеспечивает доступ к передовым образовательным ресурсам и технологиям. В умном кампусе применяются различные технологии, такие как интернет вещей (IoT), искусственный интеллект (AI), облачные вычисления и анализ данных для создания интегрированной и эффективной образовательной среды.

Цель умного кампуса – создать обучающую среду, которая адаптируется к потребностям студентов и преподавателей, обеспечивает доступ к актуальным образовательным ресурсам и технологиям, поддерживает разнообразные методы обучения и оценки успеваемости. Умный кампус предлагает инновационные методы обучения, такие как онлайн-курсы, виртуальные лаборатории и симуляции, а также разнообразные формы обратной связи на основе алгоритмов машинного обучения [1].

КОМПОНЕНТЫ УМНОГО КАМПУСА

Умный кампус включает в себя множество компонентов, каждый из которых играет важную роль в создании инновационной образовательной среды. Можно выделить следующие компоненты умного кампуса [1-3]:

- система видеонаблюдения – играет важную роль в обеспечении безопасности на кампусе. Она обеспечивает постоянное наблюдение за кампусом и используется для обнаружения и предотвращения преступлений, а также для обеспечения безопасности студентов и персонала. Система видео-

наблюдения может быть интегрирована с другими системами умного кампуса, например, с системой контроля доступа и обработки данных для обеспечения более эффективного управления физической безопасностью;

- система контроля доступа обеспечивает управляемый доступ к различным зонам на кампусе – зданиям, помещениям или ресурсам. Может быть интегрирована с системой видеонаблюдения и обработки данных для обеспечения более эффективного управления доступом в умном кампусе.

- система телекоммуникаций обеспечивает связь и обмен данными между различными устройствами и системами в умном кампусе. Используется для предоставления доступа к сети Интернет, обмена сообщениями и данными, а также для проведения онлайн-конференций и вебинаров. Может быть интегрирована с другими системами умного кампуса для обеспечения коммуникации и более эффективного обмена данными.

- система обработки данных обеспечивает анализ и обработку данных, собранных различными системами в умном кампусе. Она используется для анализа данных о безопасности, активности студентов и персонала, а также для разработки лабораторных стендов и проведения онлайн мероприятий. Интегрирована с другими системами умного кампуса для обеспечения более эффективного анализа и использования данных.

- система информационной безопасности обеспечивает защиту информации и данных, циркулирующих в умном кампусе. Она используется для обнаружения и предотвращения кибератак, защиты конфиденциальной информации и обеспечения безопасности сети. Может быть интегрирована с другими системами умного кампуса для обеспечения более эффективной защиты данных и информации.

ЛОГИЧЕСКОЕ ВЗАИМОДЕЙСТВИЕ УМНОГО КАМПУСА

Умный образовательный кампус можно представить в виде многослойной архитектуры, представленной на рис. 1:

1. Уровень датчиков – обеспечивает обогащение информацией такие системы как видеонаблюдения и контроля доступа. Также данная информация позволяет системе информационной безопасности для обнаружения и предотвращения физического вторжения.

2. Уровень доступа – обеспечивает доступ к ресурсам сети и определяет, какие пользователи или устройства имеют право получить доступ к определенным ресурсам или информационным системам умного кампуса.

3. Уровень распределения обеспечивает связь между различными сегментами и системами сети умного кампуса.

4. Уровень информационного ядра обеспечивает работоспособность бизнес-процессов, существующих в высших учебных заведениях, как то, система дистанционного обучения, управление университетом и т.д.



Структура умного образовательного кампуса

В инфраструктуре умного кампуса может применяться технология IoT, а также комплексная система информационной безопасности. Для более эффективной настройки системы информационной безопасности важно определить информационные потоки на различных уровнях логического взаимодействия умного кампуса [4].

Предложенная концепция многоуровневой модели умного кампуса является предпочтительной для решения задачи обеспечения информационной безопасности, поскольку позволяет применить подход многоэшелонированной защиты, что может предотвратить несанкционированный доступ к уровню информационного ядра умного кампуса.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Ю. И. Воротницкий, К. В. Козадаев, Е. И. Козлова, И. А. Шалатонин, Е. А. Головатая, А. М. Соболев. Цифровая интеллектуальная среда факультета // материалы II Международной научно-практической конф., Минск, 23–24 апреля 2020 г. / редкол.: В. В. Скакун (отв. ред.) [и др.]. Минск: БГУ, 2020. С.19 – 22.
2. A. Smith, B. Johnson. «Implementing Smart Campus Security Systems: Challenges and Opportunities», Journal of Smart Cities, vol. 5, no. 2, 2020. P. 123-135.
3. C. Chen, D. Wang. «Design and Implementation of a Smart Campus Security Monitoring System Based on IoT Technology», IEEE Access, vol. 7, 2019. P. 100342-100352.
4. H. Lee, S. Park. "A Review of Smart Campus Security: A Case Study Approach," // International Journal of Security and Networks, vol. 13, no. 3, 2018. P. 137-150.

РАЗРАБОТКА СИСТЕМЫ АТМОСФЕРНЫХ ЗОНДОВ ДЛЯ ОРГАНИЗАЦИИ КЛИЕНТ-СЕРВЕРНОЙ СВЯЗИ

**Е.Д. Юшкевич, И.А. Кураков, Д. А. Бореико, А.С. Ильюшенко,
В.С. Баранова, А.А. Спиридонов, Д.В. Ушаков, В.А. Саечников**

Белорусский государственный университет, Минск, Беларусь

*E-mail: rct.yushkeviED@bsu.by, rct.kurakov@bsu.by, rct.boreyko@bsu.by, rct.il-
yushenAS@bsu.by, vbaranova@bsu.by, spirid@bsu.by, ushakovdv@bsu.by,
saetchnikov@bsu.by*

В данной статье рассмотрена модель и структура спутниковой системы связи на базе группировки атмосферных зондов. Также рассмотрены конструкция, принцип работы и техническое устройство самих зондов. Актуальность разработок выражается в перспективности и многовекторности создания спутниковых группировок как направления в аэрокосмической области. Использование разработанных аппаратов для развёртывания краткосрочной группировки атмосферных зондов может быть расширено для исследования широкого спектра задач мониторинга и связи.

Ключевые слова: атмосферные зонды; спутниковые группировки; спутниковые системы связи.

ВВЕДЕНИЕ

Использование спутниковых группировок в настоящее время является одной из ведущих тенденций в развитии космических систем [1]. Их применение позволяет существенно оптимизировать сбор, обработку данных (к примеру, материалов дистанционного зондирования Земли), а также их передачу, её скорость и безопасность.

В рамках эксперимента по организации модели системы связи внутри группировки была разработана система CanSat, включающая аппарат-деплойер, конструкция которого предусматривает размещение и последующий выпуск из него ещё двух аппаратов (далее MiniSat), участвующих в обмене данными с зондом-деплойером и между собой.

КОНСТРУКЦИЯ И КОМПОНОВКА ЭЛЕКТРОНИКИ АППАРАТОВ

Конструкция аппарата и деплойера, а также их общий вид представлены на рис. 1 (а, б, в). Вследствие того, что зонд-деплойер вместе с переносимыми им аппаратами должен помещаться в цилиндр высотой 220 мм и диаметром 88 мм, что соответствует отсеку полезной нагрузки ракеты-носителя, используемой при испытаниях, корпус его представляет собой цилиндр вышеупомянутого размера.

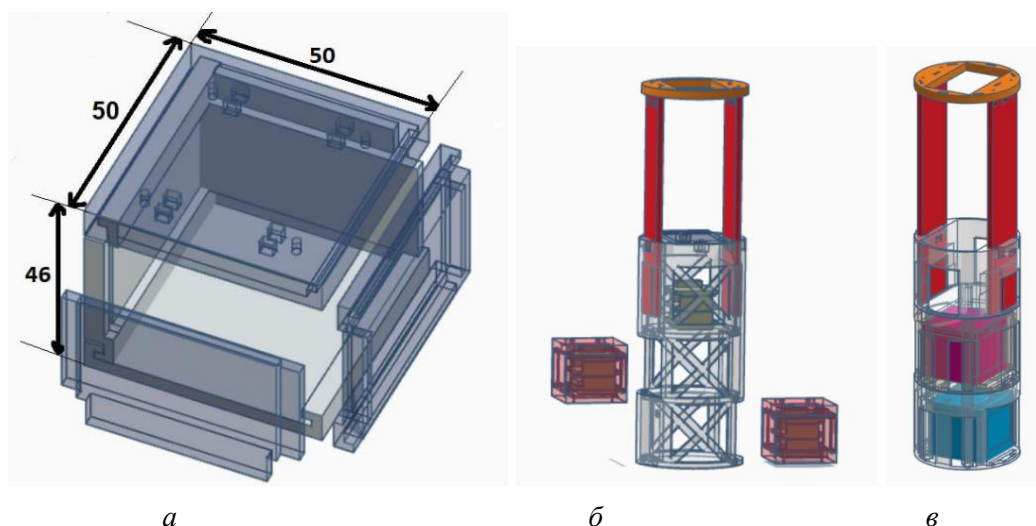


Рис. 1. Конструкция аппаратов:

а – переносимый аппарат, *б* – укомплектованный деплойер, *в* – общий вид аппаратов

В качестве выпускаемых деплойером аппаратов выступают два аппарата формата PocketCube [2] (MiniSat), структура и габариты которых отражены на рис. 1 (а). Конструкция аппарата разработана в виде легкосборного корпуса, состоящего из нескольких прямоугольных панелей с соединительными пазами для крепежа. В верхней части деплойера располагается отсек для бортовой электроники, а два выпускаемых аппарата располагаются друг над другом в нижней части, каждый в своём отсеке. Следует отметить, что конструкция предусматривает выпуск MiniSat-ов в разные стороны относительно деплойера для минимизации влияния данного процесса на ориентацию и положение деплойера в пространстве, а также во избежание возможного столкновения аппаратов.

Для максимально рационального использования места внутри аппаратов, электроника как MiniSat-ов, так и зонда-деплойера располагается на нескольких печатных платах в виде стека (рис. 2 (а, б)).

СИСТЕМА СВЯЗИ В ГРУППИРОВКЕ

В рамках проводимых в дальнейшем экспериментов зонд-деплойер, укомплектованный двумя MiniSat-ами, запускается на высоту 1500 м при помощи ракеты-носителя с удельным импульсом порядка 100 Н [3]. После выпуска из ракеты-носителя осуществляется выпуск зондов из аппарата-деплойера. Это событие фиксируется программно, после чего начинается основная фаза эксперимента, заключающаяся в непосредственном сборе и обмене данных между аппаратами.

Такая структура связи значительно упрощает взаимодействие с группировкой аппаратов, будь то отправка команд нескольким или одному её

элементу, получения собранных данных или общая техническая организация наземной станции. Однако, при наличии одного ключевого канала связи с группировкой пристального внимания требует обеспечение его безопасности.

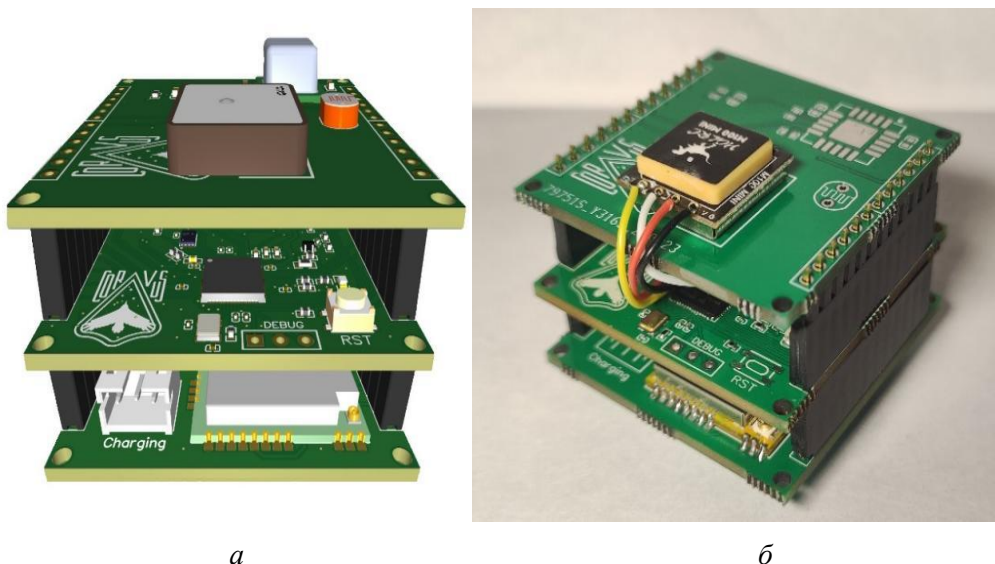


Рис. 2. Электроника зонда-деплойера:
а – модель, б – в процессе изготовления

В сформированной системе радиосвязи зонд-деплойер выполняет роль концентратора, который отвечает за организацию потока данных в системе, и весь основной трафик между группировкой зондов и наземной станцией модерируется именно им. Его задачей является сбор данных с двух других зондов, формирование содержания и очередности пакетов, отправка их на наземную станцию, а также приём и маршрутизация инструкций с наземной станции.

В данной модели группировки проблема защищённости основного канала связи решена следующим образом: в начале каждой итерации связи между группировкой и землёй MiniSat-ы, выпускаемые деплойером, генерируют две части «пароля», интегрируемого в каждый передаваемый пакет в канале деплойер-земля, что позволяет значительно повысить безопасность передачи данных. После данного события две части пароля передаются как на наземную станцию, так и на аппарат-деплойер, где комбинируются особым образом в более сложный пароль-ключ. Алгоритм взаимодействия аппарата-деплойера и наземной станции представлен на рис. 3.

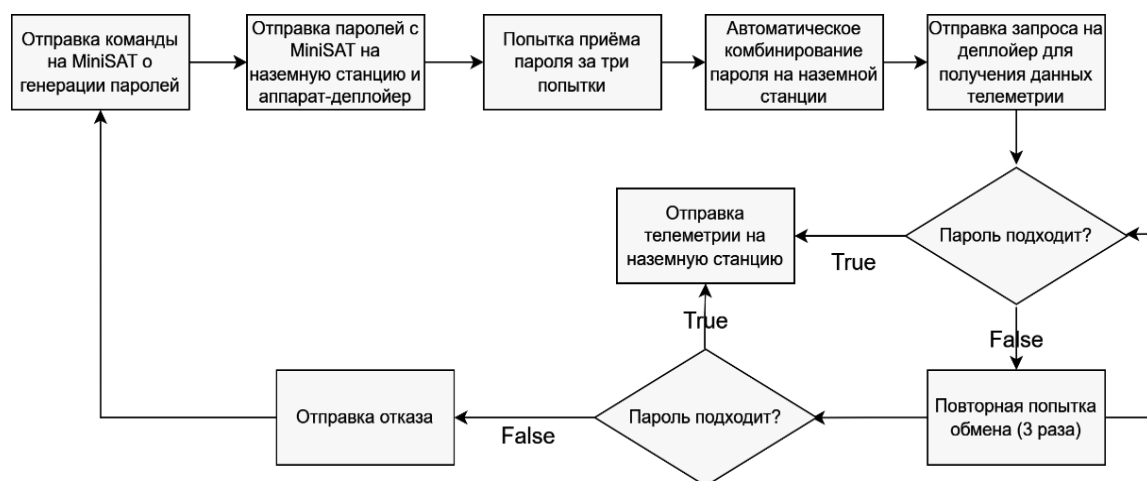


Рис. 3. Упрощённая схема взаимодействия аппарата-деплойера и наземной станции

Следует отметить, что во избежание ошибок приёма производятся три попытки отправки частей пароля. После успешного получения и обработки пароля наземная станция отправляет деплойеру пакет-запрос на получения телеметрии, что также повторяется три раза. После того, как полученный экземпляр пароля совпадает с сохранённым на зонде-деплойере, осуществляется отправка телметрии. В противном случае оправляется отказ, который инициирует начало новой итерации связи и генерации частей пароля.

Разработаны конструкция деплойера и двух аппаратов формата PocketCube для развёртывания краткосрочной группировки атмосферных зондов. Для полетного эксперимента предложена модель и структура клиентско-серверной связи между зондом-деплойером, наземной станцией и двумя аппаратами PocketCube.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. A.A. Spiridonov, V.S. Baranova, V.A. Saetchnikov, D.V. Ushakov. The regional nanosatellite constellation modelling formation by a piggyback launch from different spaceports // Journal of the Belarusian State University. Physics. 2022. № 2. P. 50-59.
2. Разработка систем космических аппаратов / редкол.: П. Фортескью, Г. Суайнерда, Д. Старка; Пер. с англ. М.: Альпина Паблишер, 2016. С. 764.
3. Cappelletti C. Cubesat Handbook // Battistini S., Malphrus B. K. Academic Press, 2021. С. 461.
4. Разработка система приема сигнала для организации трассы оптической связи в условиях атмосферы / Юшкевич Е.Д., Бореико Д.А., Гринько С.С., Ильюшенко А.С. [и др.] // RusNanoSat-2023: Сборник тезисов докладов пятого российского симпозиума по наноспутникам с международным участием. Самара: Издательство Самарского университета, 2023. С. 177-179.

РАЗРАБОТКА МОДУЛЯ ОПТИЧЕСКОЙ СВЯЗИ ДЛЯ МЕЖБОРТОВОЙ И МЕЖСПУТНИКОВОЙ КОММУНИКАЦИИ

**А.А. Ярмантович, О.Д. Гринадёр, В.С. Баранова,
А.А. Спиридонов, Д.В. Ушаков, В.А. Саечников**

Белорусский государственный университет, Минск, Беларусь

E-mail: leshayarmantovich98@gmail.com, oleggrinaderov@gmail.com, vbaranova@bsu.by, spirid@bsu.by, saetchnikov@bsu.by, ushakovdv@bsu.by

В работе рассматриваются вопросы межспутниковой и межбортовой коммуникации малых космических аппаратов. Предложена упрощённая модель оптической связи для передачи данных между наземными имитаторами наноспутника, а также их бортовыми системами. Представлены результаты передачи текстовых сообщений и изображений.

Ключевые слова: межспутниковая коммуникация; оптическая связь; протокол оптической связи NEC; защищенный протокол передачи данных.

ВВЕДЕНИЕ

Одним из ключевых аспектов развития малых космических аппаратов является обеспечение эффективной межспутниковой и межбортовой коммуникации. Беспроводная оптическая связь, в частности лазерная, обладает рядом преимуществ, делающих ее перспективным решением для CubeSat. Лазерная связь обеспечивает гигабитные скорости передачи данных, что значительно превосходит возможности традиционных радиочастотных каналов [1]. В контексте межбортовой коммуникации, лазерная связь дает дополнительные преимущества: исключение элементов электрических соединений позволяет уменьшить вес и освободить место для полезной нагрузки. Беспроводная оптическая шина обеспечивает быстрый и эффективный обмен данными между различными компонентами наноспутника. Несмотря на явные преимущества, у оптической связи есть и недостатки, которые стоит учитывать. Лазерный луч необходимо точно направить на приемник, что может быть затруднительно, особенно при наличии вибраций или других возмущений. Дальность действия оптической связи ограничена по сравнению с радиочастотной, особенно в условиях атмосферных помех [2].

В работе исследуется технология реализации оптической линии связи для межспутниковой и межбортовой коммуникации наземных учебных имитаторов наноспутника.

УПРОЩЕННАЯ МОДЕЛЬ ОПТИЧЕСКОЙ ЛИНИИ СВЯЗИ

На рис. 1 представлена структурная схема оптической системы связи, где Raspberry Pi Pico является центральным управляющим устройством и

отвечает за генерацию сигналов для ИК-светодиода (L-34F3C), а также обработку сигналов, полученных от ИК-приемника (TSOP34833). ИК-светодиод преобразует электрические сигналы в инфракрасные импульсы, приемник – импульсы света в электрические колебания, ИК-приемник преобразует инфракрасные импульсы в электрические сигналы.

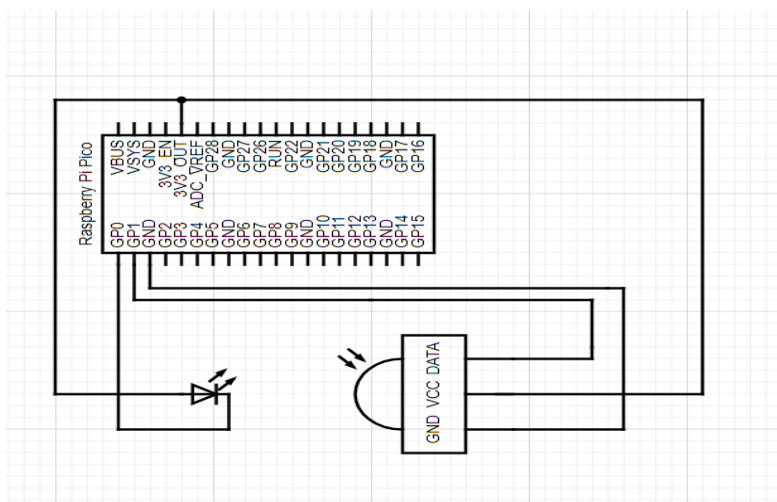


Рис. 1. Электрическая схема подключения модуля оптической связи

ПРОТОКОЛ ПЕРЕДАЧИ ДАННЫХ В ОПТИЧЕСКОМ КАНАЛЕ

Система использует модулированное инфракрасное излучение для передачи данных от ИК-светодиода к ИК-приемнику. Существует множество протоколов передачи данных для ИК-связи, и в данной работе используется протокол NEC, основанный на модуляции длительностью пауз (МДП). В протоколе NEC логические единица и ноль кодируются с помощью пауз разной длительности. Длительность импульса всегда одинакова, а значение бита определяется длительностью паузы после импульса: короткая пауза соответствует 0, а длинная – 1 [4]. Структура пакета представлена на рис. 2



Рис. 2. Структура пакета протокола NEC

ОБМЕН ДАННЫМИ В ОПТИЧЕСКОМ КАНАЛЕ СВЯЗИ

Представленная последовательность функций исходного кода реализует передачу данных по инфракрасному каналу связи с использованием протокола NEC. Алгоритм передачи данных в коде работает следующим образом:

1. *Инициализация.* Определяется GPIO-пин, к которому подключен ИК-светодиод. Инициализируется объект PWM для управления ИК-светодиодом. Устанавливается частота модуляции ИК-светодиода (33 кГц). Устанавливается скважность PWM для ИК-светодиода (50%);

2. *Функции `send_pulse()` и `send_start_pulse()`.* `End_pulse()` генерирует один импульс ИК-излучения длительностью 560 мкс. `Send_start_pulse()` генерирует стартовую последовательность протокола NEC (импульс 9 мс + пауза 4,5 мс).

3. *Подготовка данных.* Сообщение "the transfer was successful" преобразуется в массив байтов;

4. *Функция `send_package()`.* Отправляет стартовую последовательность. Циклически проходит по каждому байту в пакете данных. Для каждого бита в байте отправляет импульс ИК-излучения. Если бит равен 1, делает паузу 1200 мкс. Если бит равен 0, делает паузу 200 мкс. Делает паузу 5 мс после отправки пакета.

Алгоритм приема данных в оптическом канале:

1. *Инициализация.* Определяется GPIO-пин, к которому подключен ИК-приемник. Инициализируется объект Pin для работы с ИК-приемником.

2. *Функция `read_data()`.* Ожидает начало передачи данных (стартовая последовательность). Считывает биты данных с ИК-приемника. Преобразует биты в массив значений True/False. Возвращает массив данных.

3. *Основной цикл.* Вызывает функцию `read_data()` для считывания данных. Преобразует массив данных в строку. Выводит строку на экран. В качестве тестового сообщения отправлен и принят текст "the transfer was successful". Консоль приемника представлена на рис. 3.

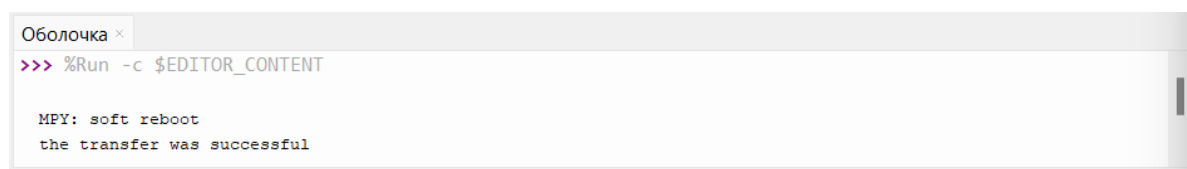


Рис. 3. Консоль приемника

Для исследования возможности передачи медиафайлов использовалось бинарное изображение размером 80x80. При передаче изображения

по инфракрасному каналу связи некоторые символы переданы некорректно, что привело к искажению отдельных пикселей, Исходное и принятое изображения представлены на рис. 4 (а, б).

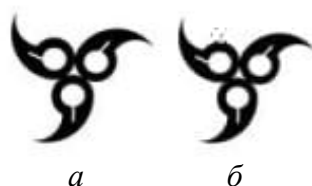


Рис. 4. а - Исходное изображение; б - принятое изображения из оптического канала связи

Скорость передачи и обработки данных составила 553,2 бит/с с битовой ошибкой 0,25%. Низкая скорость передачи и обработки данных, которая обусловлена реализацией всего алгоритма с использованием Micro Python, является возможной причиной потери данных и искажения изображения. Высокая скорость передачи данных в оптическом диапазоне не согласована с низкой скоростью обработки, реализованной на Micropython. Также, причиной искажения могли стать внешние помехи, например, солнечный свет или другие источники инфракрасного излучения.

ЗАКЛЮЧЕНИЕ

Предложена концепция реализации оптической линии связи для межспутниковой и межбортовой коммуникации наземных учебных имитаторов наноспутника на основе протокола передачи NEC. Реализован прототип упрощённой модели оптической связи для передачи данных между имитаторами наноспутника на основе программируемого контроллера Raspberry Pi Pico (центральное управляющее устройство), ИК-светодиода L-34F3C (источник оптического сигнала), ИК-приемника TSOP34833. Представлены результаты передачи текстовых сообщений и изображений. Для прототипа скорость передачи изображения составила 553,2 бит/с с битовой ошибкой 0,25%. Дальнейшие исследования предполагают использование высокоскоростных методов обработки и протоколов связи, что позволит согласовать физический канал связи с приемной частью.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Д. М. Кадочников, А. В. Кададова, Н. В. Сотникова, А. В. Трилис, В. В. Уткин Построение Систем Оптической Связи Для Малых Космических Аппаратов // Балтийский государственный технический университет «ВОЕНМЕХ» им. Д. Ф. Устинова, Санкт-Петербург, Россия. 2023
2. Impact of atmospheric turbulence on the performance of optical inter-satellite links. / Kudielka, K. [et al.] // In Proc. SPIE (Vol. 8971, pp. 89710A-1–89710A-12). 2014
3. Описание ИК-протокола NEC (кодирование длиной паузы) [Электронный ресурс]. URL: <https://radioham.ru/nec/>

РЕКОНФИГУРИРУЕМАЯ АППАРАТНО-ПРОГРАММНАЯ ПЛАТФОРМА ДЛЯ СДАЧИ НОРМАТИВОВ КОМПЛЕКСА ГТО

А.С. Семенович, С.Н. Семенович, И.П. Стецко

Белорусский государственный университет, Минск, Беларусь

E-mail: rct.semenoviAS1@bsu.by

В работе приведены результаты проектирования универсальной аппаратно-программной платформы для создания электронного тренажера с возможностью измерения времени выполнения задания, контролем качественных и количественных параметров.

Ключевые слова: электронный тренажер; облачные технологии; сбор; передача и хранение измерительных данных; оценка спортивных результатов.

В Республике Беларусь разработана концепция развития физической культуры и спорта, включая цели, задачи и принципы, способствующие повышению качества физического воспитания. Одна из важных её составляющих – проведение мониторинга физической подготовки, развития детей и молодежи, состояния взрослых. Спортивная парадигма «быстрее, выше, сильнее» стимулирует развитие инноваций как в части физических упражнений, так и спортивного оборудования. В частности, применение беспроводных ячеистых сенсорных сетей, состоящих из большого количества автономных первичных датчиков, которые могут быть размещены непосредственно на спортсмене или интегрированы в спортивный инвентарь, открывают новые возможности для сбора данных и анализа уровня физической подготовленности [1].

Авторами разработана реконфигурируемая аппаратно-программная платформа, предназначенная для оптимизации процесса сбора, мониторинга и анализа данных о выполнении физических упражнений, входящих в состав Государственного физкультурно-оздоровительного комплекса Республики Беларусь, известного под названием «Готов к труду и обороне».

Предложенная концепция заключается в создании универсальной платформы для масштабируемого тренажёра, основой которой служит автономное оконечное устройство с набором измерительных датчиков и беспроводным каналом связи [2].

Такое устройство может быть легко интегрировано в любую спортивную инфраструктуру за счет использования современных беспроводных технологий связи, таких как Bluetooth Low Energy и Wi-Fi. Объединение подобных узлов в единую сеть, обеспечит контроль взаимодействий между спортсменом и спортивным оборудованием и мониторинг физических параметров.

Структурная схема платформы представлена на рис. 1.

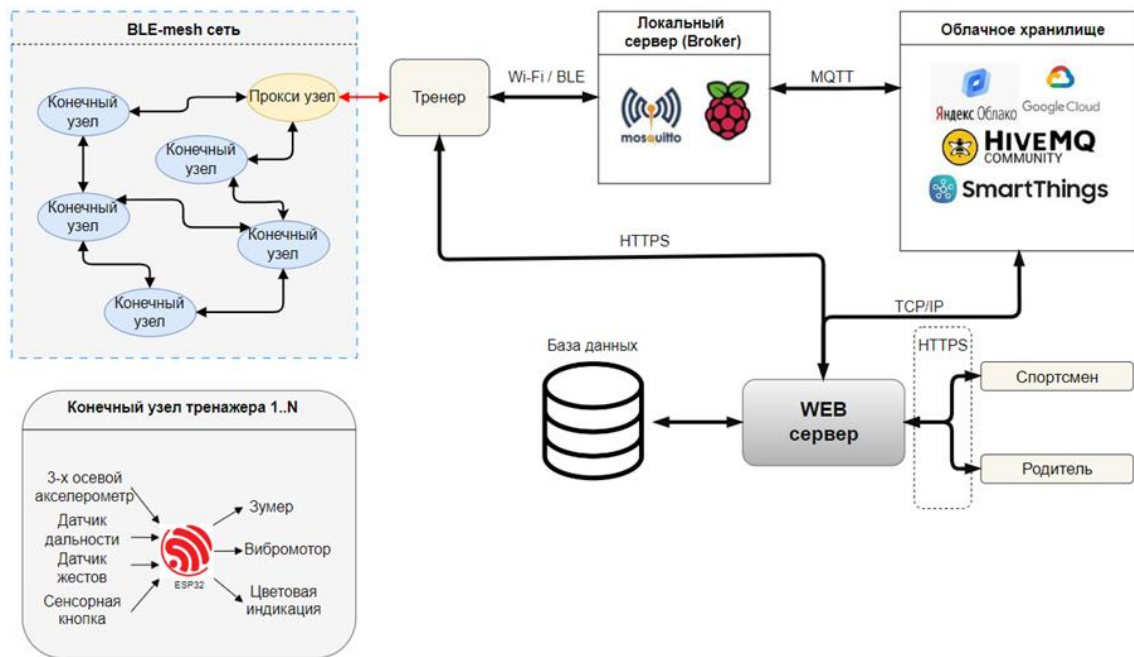


Рис. 1. Структурная схема аппаратно-программной платформы тренажера

Основная идея разработанной платформы – объединение автономных узлов-тренажеров в единую BLE-mesh сеть для создания пространственно-распределенных тренажерных комплексов, которые могут включать в себя различные стимулы и измерительные датчики обратной связи, предназначенные для мониторинга хода выполнения разнообразных физических упражнений. Синхронность регистрации данных в сенсорной беспроводной сети обеспечивается на уровне единиц миллисекунд. Это критически важный параметр для анализа времени реакции спортсмена, контроля его физических параметров и оценки эффективности тренировочного процесса [3].

В предложенной аппаратно-программной платформе используется локальный сервер, функционирующий на базе MQTT-брокера, который обеспечивает синхронизацию сигналов от распределенных датчиков (узлов), хронометрирование событий за счет применения алгоритмов IoT-технологии. Реализация такого брокера на Raspberry Pi 3 с беспроводным подключением обеспечивает автономность системы в рамках спортивного зала/стадиона и реализует возможность оперативного подключения смарт-устройств тренера или спортсмена для мониторинга тренировочного процесса в реальном режиме времени.

С помощью базового набора датчиков, интегрированных в узловые устройства платформы, можно точно фиксировать время, количество и качество выполнения ряда упражнений комплекса ГТО. К таким упражне-

ниям относятся, например: прыжок в длину с места, бег, прыжки, подтягивание на перекладине, сгибание и разгибание рук в упоре лежа, челночный бег, наклон вперед из положения сидя и т.д.

Узловое устройство платформы состоит из набора датчиков, основными из которых являются цифровой микромеханический трех-осевой акселерометр MPU6050, регистрирующий статические и динамические ускорения, и Time-of-Flight дальномер VL53L0X для определения расстояния и положения частей тела спортсмена в пространстве относительно конечным узлом тренажера.

Для выполнения функциональных тестов платформы были созданы пять универсальных электронных макета узловых устройств (рис. 2).



Рис. 2. Макеты узловых устройств:

а - 3D модель устройства; *б* - реальный образец узлового устройства; *в* - устройство в активированном состоянии

Тестирование проводилось в следующих режимах: подсчёт количества подтягиваний и прыжков на скакалке, фиксация глубины наклона вперед и локальная группа узловых устройств для проведения упражнения «челночный бег».

Группа узлов (стимул и регистратор событий) конфигурировалась для каждого вида упражнений. Результаты передавались на локальный брокер в виде структурированных JSON-сообщений. Накопленные данные анализировались и визуализировались на экране смарт-устройств.

На рис. 3 представлен график изменения виброскорости конечного узла тренажера закреплённого на поясе спортсмена в рамках выполнения упражнения «прыжки на скакалке» (слева) и «прыжок в высоту» (справа).

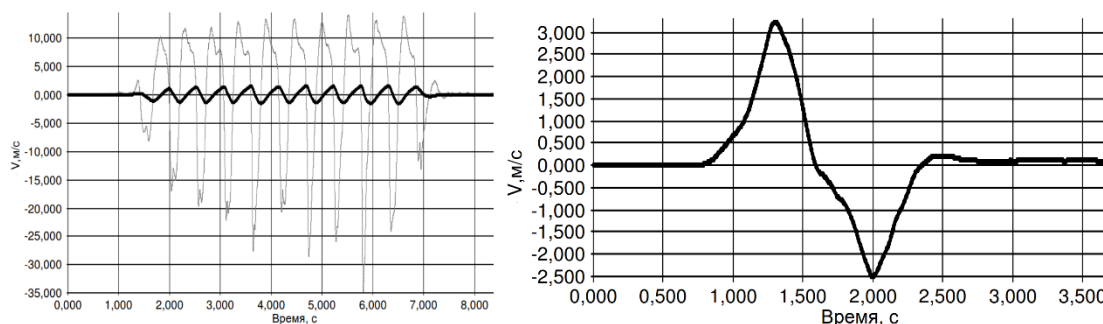


Рис. 3. Результаты исследования показаний акселерометра, закреплённого на поясе спортсмена

Режим «челночный бег» представляет комплексную тренировочную задачу, включающую использование двух или более конечных узлов, объединённых в пространственно-распределённый тренажёр. Основная цель данного режима заключается в последовательной активации цветowych стимулов на узлах сети и точной фиксации времени, необходимого спортсмену для преодоления расстояния между этими узлами.

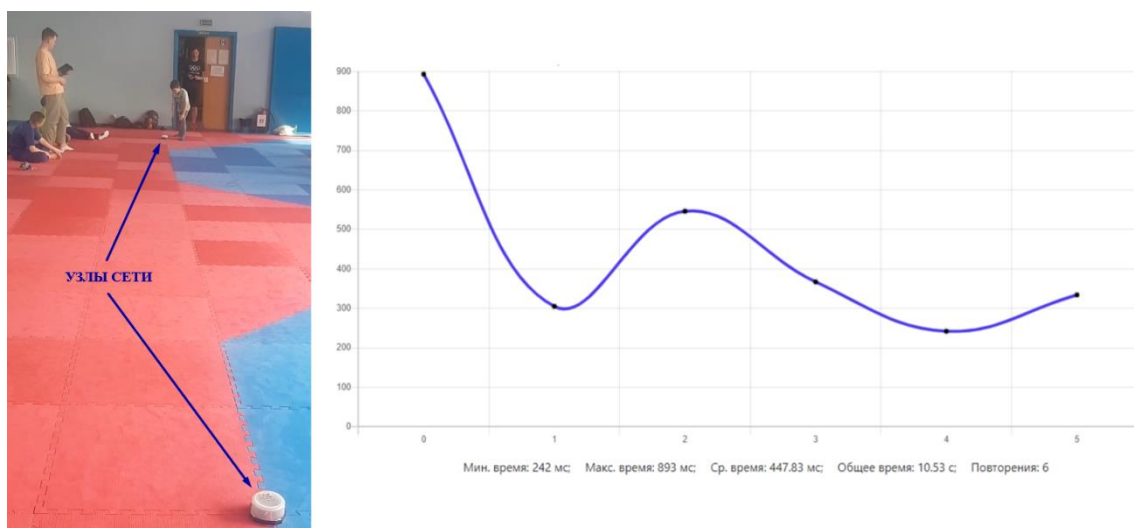


Рис. 4. Скриншот приложения для анализа выполнения упражнения «челночный бег»

ВЫВОД

Приведенные результаты тестирования созданной платформы с использованием IoT-технологий подтверждают возможность ее использования в масштабируемых задачах сбора и анализа тренировочных данных. Например, при условии создания единой базы данных, одна из возможных областей применения тренажёра в спортивной практике и физическом воспитании – республиканский мониторинг физических параметров молодежи.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Lei Liu, Xuefeng Zhang.* A Focused Review on the Flexible Wearable Sensors for Sports: From Kinematics to Physiologies. *Micromachines* 13(8):1356. 2022.
2. *A. S. M. Kayes, M. A. Hossain, M. S. Hossain, M. S. Alam, M. S. Anwar.* Wireless Sensor Networks for Sports Performance Monitoring: A Comprehensive Review. *Applied System Innovation* 1(4):52. 2018.
3. *Семенович А.С.* Масштабируемый IoT тренажер спортсмена. Сборник научных работ студентов Республики Беларусь «НИРС 2020». 2021. С 143-144.

УПРАВЛЕНИЕ ГРУППИРОВКОЙ МАЛОРАЗМЕРНЫХ КОСМИЧЕСКИХ АППАРАТОВ РАСПРЕДЕЛЕННОЙ СЕТЬЮ НАЗЕМНЫХ СТАНЦИЙ

**А.А. Спиридонов, В.С. Баранова, В.Е Евчик, И. А. Шалатонин,
В.Е. Черный, Д.В. Ушаков, В.А. Саечников**

*Белорусский государственный университет, Минск, Беларусь
E-mail: Spirid@bsu.by*

В работе представлен алгоритм распределения задач управления группировкой малоразмерных космических аппаратов между станциями наземной сети в условиях видимости максимального числа спутников. Рассмотрена модель распределения управления спутников группировки Planet между тремя наземными станциями приема Минск, Калуга, Самара.

Ключевые слова: группировка малоразмерных космических аппаратов; наземные станции приема.

Среди малоразмерных космических аппаратов (МКА)- наиболее перспективными и привлекательными с точки зрения качество, цена, время разработки, надежность функционирования является проектирование и использование группировок нано- и пикоспутников, позволяющее существенно повысить соотношение качество/эффективность для решения различных задач (радио и оптической связи, дистанционного зондирования Земли, исследование погодных явлений, слежения за подвижными объектами, научных исследований) [1]. Необходимость большого количества спутников в группировке диктуется возрастающими требованиями к непрерывному доступу к данным, которые поставляют космические системы. В идеале любая космическая услуга должна быть предоставлена в любое время, в любом месте, с требуемым качеством, за умеренную плату. На начало 2024 г. продолжали разворачиваться более 70 группировок МКА. Наземный сегмент является критически важной частью распределения задач управления и отслеживания группировок, а также сбора и распределения данных, особенно в условиях видимости максимального числа МКА, что соответствует режиму перегруженности наземных станций.

При управлении группировкой МКА решаются следующие задачи: прогнозирование пролётов МКА для организации устойчивой командной и телеметрической радиолиний; удалённое командное управление отдельными наземными станциями приема (НСП); анализ оперативной обстановки состояния МКА, оценка состояния работоспособности спутников группировки, выявление возникновения внештатных ситуаций; планирование сеансов управления и приема телеметрии со спутниками группировки; планирование внешнетраекторных измерений; планирование сценариев целевой миссии группировки [2].

Распределение МКА между НСП для решения задач по управлению группировкой в условиях видимости максимального числа спутников проводится согласно следующему алгоритму:

1. Загрузка начальных орбитальных параметров МКА группировки в формате TLE или в виде вектора состояния. Определение текущих орбитальных параметров группировки и расчет начального вектора состояния (в необходимом формате, например, в модели кругового движения используется орбитальные параметры: наклонение, период, долгота восходящего узла, аргумент широты) на интервале моделирования.

2. Расчет параметров приема каждого МКА группировки для НСП на 24 ч интервале: интервалы видимости МКА (при минимальном угле места); временная диаграмма количество видимых МКА для НСП; суммарное время видимости МКА (при минимальном угле места) для НСП.

3. Оценка временных интервалов с максимальным количеством видимых МКА для распределенного управления НСП группировкой.

4. Расчет параметров приема каждого МКА группировки для НСП на временных интервалах с максимальным количеством видимых МКА.

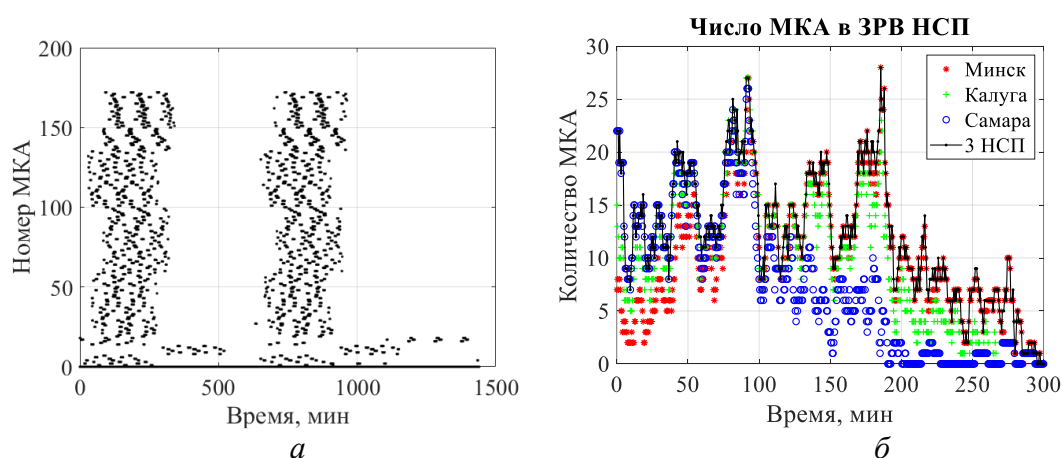
5. Разбиение интервалов с максимальным количеством видимых МКА на подинтервалы, содержащие локальные максимумы числа спутников; расчет параметров наведения антенных систем НСП и радиосигналов для каждого МКА группировки; отбор МКА с временем радиовидимости не менее 4 мин для назначения НСП на управление с учетом минимальных отклонений по параметрам наведения антенных систем и текущему приоритету (определяемому на основе анализа актуальной телеметрии и необходимостью работы с полезной нагрузкой или основными бортовыми системами).

Рассмотрим моделирование распределения управления МКА группировки ДЗЗ Planet между тремя НСП (Минск, Калуга, Самара). Группировка ДЗЗ Planet из 172 наноспутников (массой 4,7 кг и размерами 10*10*30 см), расположенных на околокруговых (эксцентриситет $e \ll 1$) солнечно-синхронных орбитах, позволяющим обеспечить постоянство времени пересечения экватора 9:30 – 11:30 (UTC). За один день спутники снимают больше 200 млн кв. км и покрывают почти весь мир самыми актуальными снимками.

Первоначально загружаются начальные орбитальные параметры МКА группировки Planet в формате TLE или в виде вектора состояния. Проводится расчет текущих орбитальных параметров группировки и расчет начального вектора состояния на момент начала времени моделирования. Как показали результаты моделирования текущих орбитальных параметров группировки Planet на 4 марта 2024 г. у большинства МКА группировки высоты орбит составляют 400-540 км, наклонения -97,02-97,50°,

при периодах обращения от 92,91 до 95,74 мин. Две основные орбитальные плоскости сдвинуты по долготе восходящего узла на угол $10\text{--}14^\circ$ при равномерном распределении МКА по аргументу широты. У 100% МКА эксцентриситет орбиты e не превосходит 0,005, что позволяет при моделировании движения спутников на кратковременных интервалах (в несколько часов) использовать модели кругового движения.

Затем проводится моделирование параметров приема каждого МКА группировки на 24 ч интервале над каждой НСП, как показано на рис. 1, а, а для основной НСП Минск. Моделирование позволяет для НСП, которые решают задачи по управлению МКА группировки, провести оценку интервалов видимости МКА, суммарного времени видимости каждого МКА, временных интервалов с максимальным количеством видимых МКА, параметров приема, и затем распределить задачи по управлению каждым МКА между НСП. Как показали результаты моделирования, за сутки пролета 06:00:00 4 марта - 06:00:00 5 марта 2024 г. над НСП Минск наблюдалось два интервала времени порядка 300 мин с временным промежутком между ними порядка 360 мин (от 06:44:00 до 11:28:00 и от 17:17:00 до 21:58:00), когда в зоне радиовидимости находилось максимальное количество видимых МКА от 5 до 28 спутников. При этом каждый МКА группировки за сутки находился в зоне радиовидимости НСП Минск (при минимальном угле места 5°) суммарно от 30 до 45 мин с максимальным интервалом на одном пролете до 8 мин.



Интервалы видимости МКА (при минимальном угле места 5°) за сутки пролета 06:00:00 4 марта - 06:00:00 5 марта 2024 г. над НСП Минск (а) и количество видимых МКА на 5 часовом временном интервале от 06:44:00 до 11:44:00 4 марта 2024 г. над НСП Минск, Калуга, Самара (б)

После этого на 5 часовых временных интервалах от 06:44:00 до 11:44:00 и от 17:17:00 до 22:17:00 4 марта 2024 г., когда над основной НСП

Минск наблюдалось максимальное количество видимых МКА, проводилось моделирование группировки для НСП Минск, Калуга, Самара, как показано на *Рис. 1, б*. На заданном интервале количество видимых МКА изменялось от 0 до 28, каждый МКА группировки находился в зоне радиовидимости (ЗРВ) трех НСП (при минимальном угле места 5°) суммарно от 7 до 29 мин с максимальным интервалом на одном пролете до 8 мин. При этом интервалы между минимумами и максимумами графика количества видимых МКА для НСП для заданного интервала от 06:44:00 до 11:44:00 меняются от 20 до 50 мин. Разбивая заданный интервал от 06:44:00 до 11:44:00 на 10 временных подинтервалов длительностью от 20 мин до 50 мин, проводится расчет параметров наведения антенных систем для НСП (угла места азимута) и принимаемого радиосигнала (доплеровского сдвига частоты) каждым МКА группировки и отбор МКА с временем радиовидимости не менее 4 мин для назначения НСП на управление с учетом минимальных отклонений по параметрам наведения антенных систем и текущему приоритету (определяемому на основе анализа актуальной телеметрии и необходимостью работы с полезной нагрузкой).

Реализован программный модуль распределения задач управления МКА группировки ДЗЗ Planet между тремя НСП (Минск, Калуга, Самара), который позволяет выявить интервалы видимости максимального числа спутников над основной станцией. Затем, определить на этих интервалах подинтервалы с общим максимальным количеством видимых спутников над всеми НСП, провести расчет параметров наведения антенных систем для НСП и принимаемого радиосигнала каждым МКА группировки. И окончательно реализовать отбор МКА с временем радиовидимости не менее 4 мин для назначения НСП управления, с учетом минимальных отклонений по параметрам наведения антенных систем и текущему приоритету, определяемому на основе анализа актуальной телеметрии и необходимостью работы с полезной нагрузкой.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. A.A. Spiridonov, V.S. Baranova, V.A. Saetchnikov, D.V. Ushakov. The regional nanosatellite constellation modelling formation by a piggyback launch from different spaceports // Journal of the Belarusian State University. Physics. 2022. № 2. P. 50-59.
2. Моделирование построения региональной группировки наноспутников попутным запуском / А.А. Спиридонов, В.С. Баранова, В.Е. Евчик, И.А. Шалатонин [и др.] // Компьютерные технологии и анализ данных (СТДА'2022): материалы III Междунар. науч.-практич. конф., Минск, 21-22 апр. 2022 г. / БГУ; редкол.: В.В. Скакун (отв. ред.) [и др.]. Минск, 2022. С. 240-243.

НЕЙРОСЕТЕВОЕ РАСПОЗНАВАНИЕ ЭМОЦИЙ НА СТАТИЧНЫХ ИЗОБРАЖЕНИЯХ

П.А. Кочеткова, Е.И. Козлова

*Белорусский государственный университет, Минск, Беларусь
kochetkova1303@gmail.com*

Представлена уточненная модель выражения эмоций на лице человека для их распознавания по изображению. Для реализации процесса распознавания использована сверточная нейронная сеть. Целевой переменной является эмоция, которая может принимать одно из значений удивление, счастье, гнев, печаль. Для увеличения количества изображений в обучающей выборке применено аугментирование данных наложением гауссовского шума. Исследования показали, что после уточнения модели количество правильно распознанных эмоций на изображениях повысилось на 5 - 12 % в зависимости от типа эмоции.

Ключевые слова: распознавание эмоций; сверточная нейронная сеть; набор данных; аугментация.

ВВЕДЕНИЕ

Распознавание эмоций по изображениям лица человека является сегодня актуальной задачей при создании приложений и алгоритмов управления в различных областях человеческой деятельности – рекламном деле, социальной и медицинской сферах, образовании, технических областях, на транспорте, в развлекательном бизнесе и пр. В последнее время эта задача решается с помощью нейросетевых алгоритмов [1]. В ряде работ исследователей нейросетевых методов распознавания эмоций [2–4] описаны разные архитектуры свёрточных нейронных сетей, применяемые для распознавания эмоций на изображениях, полученных из видеопотока. Чаще всего исследователи применяют сверточные нейронные сети как ставшие уже классическими, так и с изменением числа слоев, функций активации и пр. Так, авторы [2] указали, что при использовании двух разных архитектур после обучения полученная в первом случае точность распознавания эмоций составила всего 52 % на тестовом наборе данных, а после изменения архитектуры – 64 %. При этом на обучающем наборе данных точность распознавания составила в обоих случаях около 98 %. В данной работе исследованы возможности повышения качества распознавания эмоций путем уточнения моделей описания некоторых базовых эмоций человека и применения сверточной нейронной сети для их распознавания в соответствии с уточненными моделями.

ИНСТРУМЕНТЫ ИССЛЕДОВАНИЯ

Сверточные нейронные сети (CNN) схематично представляют собой последовательность слоев, каждый из которых преобразует один активационный объем в другой с помощью дифференцируемой функции. Для организации сверточной нейронной сети применяется 3 основных слоя: свертка, пулинг и полносвязный слой. Слои свертки и пулинга используются для извлечения карты признаков из исходного изображения, а полносвязные слои используются для конечной классификации изображения по извлеченным признакам [5].

Для проведения исследований по распознаванию эмоций на изображениях лиц в качестве инструмента были выбраны свёрточные нейронные сети (CNN). Выбранная архитектура CNN содержит 2 слоя свертки, 2 pooling-слоя и 4 полносвязных слоя. Архитектура подобрана так, чтобы результаты можно было сравнить с результатами [2]. Отличием был размер входного слоя сети (он равнялся $256 \times 256 \times 1$), определенный в соответствии с размером изображений из набора данных. Выходной слой сети – это вектор из 4-х компонентов со значениями от 0 до 1, указывающими степень принадлежности предъявленной на изображении эмоции к выбранному классу. В данном исследовании также применена функция активации ReLU. Для обучения применен метод обучения с учителем, при котором сеть обучается на большом наборе аннотированных изображений лиц с различными эмоциями. Для улучшения производительности и предотвращения переобучения применен метод градиентного спуска.

УТОЧНЕНИЕ МОДЕЛЕЙ БАЗОВЫХ ЭМОЦИЙ И ПОДГОТОВКА ИЗОБРАЖЕНИЙ

К базовым эмоциям принято относить печаль, гнев, удивление, страх, отвращение, презрение, радость (эмоции удовольствия). В данной работе рассматривались следующие эмоции: удивление, счастье (радость), гнев, печаль [6].

Для обучения нейронной сети необходимо иметь большой набор данных, содержащий изображения лиц с различными эмоциями. В качестве основного набора данных для исследований был использован "Facial Emotion Recognition Image Dataset" [7]. Выбранный датасет содержит изображения людей разных возрастов, пола, расы, разной степени экспрессии выражения эмоций.

Анализ результатов [2] позволяет сделать предположение, что высокая точность распознавания на обучающей выборке и относительно низкая точность распознавания на тестовом наборе являются признаком переобучения сети. При этом регуляризация совместно с добавлением новых сверточных слоев в модели улучшила точность распознавания на 12% [2].

Мы предположили, что изменение моделей базовых эмоций и отбор изображений для обучения нейронной сети согласно уточненным признакам модели, позволит улучшить качество распознавания эмоций по изображению лица человека. Для проверки этого предположения и с целью уменьшить неоднозначность в выражении эмоций на снимках на этапе обучения нейронной сети для повышения качества распознавания, были сделаны уточнения к моделям, описанным в [6]. В дальнейшем отбор изображений и их разметка для обучения сети проводились в соответствии с приведенными ниже описаниями выражения эмоций.

Удивление: брови округлены и высоко подняты; горизонтальные морщины пересекают весь лоб; верхние веки подняты, а нижние – опущены; напряжения или натяжения в области рта нет, челюсть опущена так, что рот открыт, губы и зубы разъединены.

Счастье: уголки губ оттянуты в стороны и приподняты; рот может быть приоткрыт или закрыт, а зубы обнажены или прикрыты губами; от носа к внешнему краю губ (позади уголков губ) тянутся морщины – носогубные складки; щеки подняты; под нижними веками образуются морщинки, нижние веки могут быть приподняты, но не напряжены; у наружного края уголков глаз появляются морщинки («гусиные лапки»).

Гнев: брови опущены и сведены; между бровями появляются вертикальные складки; нижние веки напряжены и могут быть как подняты, так и оставаться без изменений; верхние веки напряжены и могут не изменить своего положения или же опуститься вследствие движения бровей; глаза неподвижны и могут казаться выпученными; губы находятся в одном из двух положений: а) крепко сжаты, причем уголки губ могут быть прямыми или опущенными или же б) раскрыты и напряжены в форме квадрата, как при крике; ноздри могут быть расширены.

Печаль: изогнутые вниз губы и поднимающиеся внутрь брови с образованием морщин на лбу; верхние веки слегка опущены; рассеянный взгляд, уголки рта слегка опущены; внутренние уголки бровей подняты вверх; кожа под бровями становится треугольной формы; внутренние уголки верхних век приподняты; уголки губ опущены или же губы расслаблены.

Изображения были предварительно масштабированы и обрезаны, а также переведены в оттенки серого. Для увеличения объема и разнообразия данных была проведена аугментация, в частности, к имеющимся снимкам был добавлен гауссовский шум в разных процентных отношениях к исходному изображению. Таким образом удалось увеличить датасет до 10000 изображений. Далее данные были разделены на выборки – обучающую, валидационную и тестовую в соотношении 70-15-15 %.

РЕЗУЛЬТАТЫ И ВЫВОДЫ

В результате проведенной работы установлено, что уточнение модели приводит к повышению числа правильно распознанных эмоций из тестовой выборки на 5-12 % в зависимости от вида эмоции. При этом на обучающей выборке результат достигал, как и в случае, описанном авторами [2], порядка 98%, однако на тестовой и валидационной – не превышал 79 и 64 % соответственно. Этот результат нельзя назвать отличным, однако он указывает на возможность повышения качества распознавания эмоций на основе комплексного подхода – уточнения модели эмоций, изменения архитектуры сети и расширения датасета методами аугментации и другими. В дальнейшем планируется расширить набор данных для обучения сети, применить различные архитектуры, а также функции активации, провести тестирование её работы на изображениях с несколькими лицами, а также с кадрами из видеопотока.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Бергер Е. Г., Бублик И. Ю. Анализ применения искусственного интеллекта на базе технологии отслеживания эмоций человека // *Colloquium-journal*. 2020. №10 (62).
2. Тимофеева Ольга Павловна, Неимущев Сергей Андреевич, Неимущева Любовь Ивановна, Тихонов Иван Александрович. Распознавание эмоций по изображению лица на основе глубоких нейронных сетей // *Труды НГТУ им. Р. Е. Алексеева*. 2020. №1 (128).
3. Рюмина Е.В., Карпов А.А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // *Научно-технический вестник информационных технологий, механики и оптики*. 2020. №2.
4. Нейросетевой алгоритм распознавания эмоций человека по изображению лица [Электронный ресурс]. URL: <https://elib.bsu.by/bitstream/123456789/248683/1/245-248.pdf> (дата обращения 30.03.2024)
5. Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество [Электронный ресурс]. URL: <https://habr.com/ru/articles/348000/> (дата обращения 18.02.2024)
6. Экман П. Психология эмоций. Я знаю, что ты чувствуешь//изд. Питер; СПб. 2010. С. 235. ISBN 5–49807–705–5
7. Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor. "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," in *IEEE Transactions on Affective Computing*, vol. 10, no. 1, 1 January-March 2019. P. 18-31. DOI: 10.1109/TAFFC.2017.2740923.

ИССЛЕДОВАНИЕ ВЗАИМОСВЯЗИ МЕЖДУ ВАРИАБЕЛЬНОСТЬЮ СЕРДЕЧНОГО РИТМА И ВЫРАЖЕННОСТЬЮ РЕАКЦИИ ВЕСТИБУЛЯРНОГО АППАРАТА ЧЕЛОВЕКА ПРИ ПОГРУЖЕНИИ В СРЕДУ ВИРТУАЛЬНОЙ РЕАЛЬНОСТИ

Шан Вэньли, Е.И. Козлова

*Белорусский государственный университет, Минск, Беларусь
wenlishang128@gmail.com*

В данной работе исследовалось наличие и выраженность связи между вариабельностью сердечного ритма (BCP) и степенью эффекта укачивания при нахождении человека в сценах виртуальной реальности (VR-MS), а также особенности этой связи. Для прогнозирования вариабельности сердечного ритма и VR-MS были использованы модели машинного обучения. Определено, что модель MLPRegressor наилучшим образом соответствует результатам экспериментов и может быть применена для прогнозирования характера связи между исследованными переменными.

Ключевые слова: виртуальная реальность; укачивание; прогнозирование состояния; машинное обучение; VR сценарий.

ВВЕДЕНИЕ

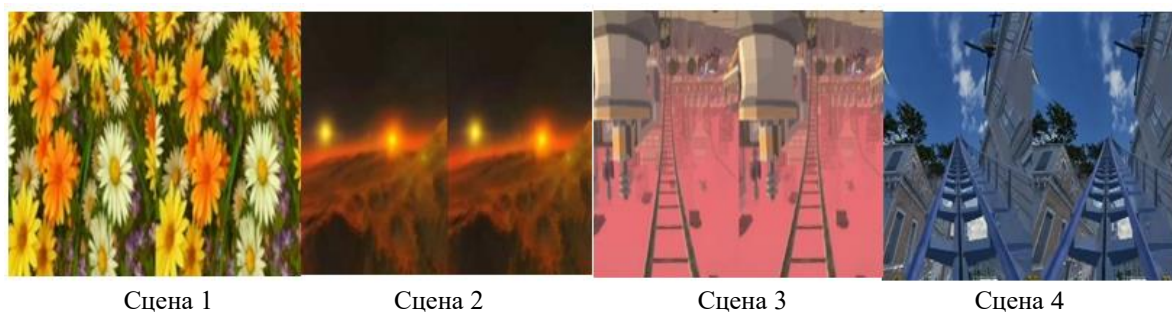
В последние годы технология виртуальной реальности (VR) привлекает все большее внимание благодаря своим иммерсивным интерактивным возможностям [1 – 3]. Исследование влияния VR-окружения на физиологические реакции, например, вариабельность сердечного ритма, субъективные переживания, такие как движение виртуальной реальности, является одной из областей, которая привлекает большой интерес [4–8].

Понимание связи между этими факторами важно для улучшения пользовательского опыта и решения потенциальных проблем здоровья в приложениях VR. В этой статье исследована связь между вариабельностью сердечного ритма и степенью тяжести укачивания в различных сценариях VR, при этом учитываются факторы окружения и индивидуальные различия испытуемых.

МАТЕРИАЛЫ И МЕТОДЫ ИССЛЕДОВАНИЯ.

Экспериментальное проектирование. В рамках данного исследования испытуемые погружались в различные сценарии VR, одновременно измерялись их физиологические реакции и субъективные переживания. В исследовании участвовало 35 участников, которые имели различный уровень опыта взаимодействия с технологией VR. Были разработаны шесть различных сценариев VR, каждый из которых отличался по содержанию и интенсивности движения. Сценарии включали в себя имитацию американских горок, виртуальный туризм и интерактивные игровые среды.

Сцены: Виртуальные сцены представлены в пассивном движении: испытуемый остается неподвижным в первом лице, в то время как сцена представляется ему в виде пассивного движения. Сцены от первой до третьей: использование трех различных типов и скоростей виртуальных движущихся сцен. Сцены от четвертой до шестой: использование одной и той же сцены виртуальной реальности, но с разной скоростью воспроизведения, чтобы сравнить влияние различных скоростей воспроизведения на испытуемых. В частности, скорость воспроизведения для сцены четвертой составляет 2,0, для сцены пятой - 1,0, для сцены шестой - 0,5.



Кадры VR-сцен, использованных в эксперименте

Физиологические данные (включая вариабельность сердечного ритма) собирались с помощью носимых биоинформационных датчиков. Субъективные оценки степени тяжести укачивания получались после каждого опыта с VR с помощью опросника.

Оборудование: VR-очки для предоставления пользователю виртуального опыта, смартфон REALME gt5PRO для воспроизведения виртуальных движущихся сцен, и браслет мониторинга здоровья DIDO g28SPRO для отслеживания физиологических показателей испытуемых, таких как сердечный ритм.

Анализ данных. Для изучения связи между вариабельностью сердечного ритма и степенью тяжести движения в различных сценариях VR, а также их различий в зависимости от сценария, использовались методы статистического анализа, включая дисперсионный анализ и коэффициенты корреляции Пирсона и Кендалла. С использованием моделей машинного обучения предсказывалась степень тяжести укачивания на основе данных о вариабельности сердечного ритма и результатов опросов участников эксперимента. Учитывая особенности данных, мы применили несколько моделей машинного обучения, включая Support Vector Machine, Random Forest, Gaussian Regression, Decision Tree, Kernel Ridge, Integrated Models, MLPRegressor и SequentialModel. Эти модели были применены для подгонки данных по BCP

и оценкам VR-укачивания, а также оценили их производительность на тестовом и валидационном наборах данных. Учитывая производительность и сложность моделей, мы выбрали MLPRegressor в качестве окончательной модели для прогнозирования. Эта модель показала отличные результаты на тестовом и валидационном наборах данных, имея низкое MSE и высокий R-squared, что свидетельствует о ее высокой точности и способности к обобщению. Проведенный анализ и сравнение производительности других моделей показал, что несмотря на хорошие результаты на тестовом наборе данных, их производительность на валидационном наборе была низкой, что может свидетельствовать о переобучении или недостаточной способности к обобщению.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ.

Во всех VR-сценариях наблюдалась умеренная положительная корреляция между вариабельностью сердечного ритма и степенью тяжести укачивания. Однако сила этой связи варьировалась в зависимости от конкретных характеристик каждого сценария. Модель MLPRegressor успешно предсказывала степень тяжести укачивания на основе данных о вариабельности сердечного ритма, имея низкую среднеквадратическую ошибку и высокий коэффициент детерминации R.

Контекстуальные факторы. Результаты данного исследования свидетельствуют о том, что контекстуальные факторы, такие как содержание VR-сцены и интенсивность движения, играют важную роль во влиянии на связь между физиологическими реакциями и степенью тяжести укачивания. Сцены с более высокой интенсивностью движения, как правило, вызывали более сильные физиологические реакции и более высокие оценки степени укачивания. Так, анализ значений коэффициента корреляции Кендалла между ИСР и оценками ВР-укачивания показал относительно высокие значения (0,59, 0,61 и 0,48) в сценах с первой по третью, при этом соответствующие значения уровня значимости (p) также были очень низкими: 2,07e-05, 3,08e-06 и 0,00017. Это свидетельствует о том, что в каждой сцене существует сильная упорядоченная связь между ИСР и оценками ВР-укачивания, а также наличие высокой статистической значимости этих корреляций, а не случайность.

Индивидуальные различия. В реакции участников на VR-сцены наблюдались индивидуальные различия, обусловленные не только личной восприимчивостью к укачиванию, например, в транспорте, но и различием в сценариях и скорости воспроизведения представленных испытуемым сцен. Эти индивидуальные различия подчеркивают важность индивидуального подхода к разработке и применению VR.

ВЫВОДЫ

Относительно небольшой объем выборки может ограничивать обобщение результатов исследования. Кроме того, индивидуальные различия и субъективные факторы в ходе эксперимента могут влиять на стабильность результатов. В связи с этим, в будущем следует расширить объем выборки и провести дополнительное изучение влияния индивидуальных характеристик, эмоционального состояния и других факторов на ИСР и оценки ВР-укачивания. Предполагается также провести дополнительную оптимизацию выбора моделей и методов анализа данных для улучшения научной обоснованности и надежности исследования. Проведенные исследования помогают нам понять, как физиологические реакции и субъективные переживания влияют на восприятие информации в среде виртуальной реальности. Учитывая факторы окружения и индивидуальные различия, мы можем лучше разрабатывать виртуальные среды, чтобы максимально снизить дискомфорт от движения и повысить комфорт и удовлетворение пользователя. Дальнейшие исследования в этой области представят дополнительную информацию для разработки более погруженных и пользовательски-ориентированных приложений VR.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Arcioni B., Palmisano S., Apthorp D., Kim J.. Postural stability predicts the likelihood of cybersickness in active HMD-based virtual reality. *Displays* 58: P. 3–11. 2019. DOI: <https://doi.org/10.1016/j.displa.2018.07.001>.
2. Kim J, Chung C.Y.L., Nakamura S., Palmisano S., Khuu S.K.. The Oculus Rift: a cost-effective tool for studying visual-vestibular interactions in self-motion perception. *Front Psychol* 6: P. 248. 2015. DOI: <https://doi.org/10.3389/fpsyg.2015.00248>
3. Stanney K.M., Hale K.S., Nahmens I, Kennedy R.S. What to expect from immersive virtual environment exposure: influences of gender, body mass index, and past experience. *Hum Factors* 45(3): P. 504–520. DOI: <https://doi.org/10.1518/hfes.45.3.504.27254>
4. Reason J.T., Brand J.J. Motion sickness. Academic Press, New York. 1975.
5. Riccio G.E., Stoffregen T.A. An ecological theory of motion sickness and postural instability. *Ecol Psychol* 3(3): 1991. P. 195–240. DOI: https://doi.org/10.1207/s15326969eco0303_2
6. Ebenholtz S.M., Cohen M.M., Linder B.J. The possible role of nystagmus in motion sickness: a hypothesis[J]. *Aviation, space, and environmental medicine*, 1994. 65(11): P. 1032-1035.
7. Holmes S.R., Griffin M.J. Correlation between heart rate and the severity of motion sickness caused by optokinetic stimulation[J]. *Journal of Psychophysiology*, 2001. 15(1): P. 35.
8. Tokatli A., Durmus C. Correlation between heart rate variability and motion sickness caused by optokinetic stimulation[J]. *Balkan Military Medical Review*. 2014. P. 17.

СЕКЦИЯ «КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ПРОЦЕССОВ И СИСТЕМ»

МОДЕЛИРОВАНИЕ ФОТОТОКА В КРЕМНИЕВЫХ ФОТОДИОДАХ С PIN-СТРУКТУРОЙ

А.В. Борздов¹⁾, В.М. Борздов¹⁾, А.Н. Петлицкий²⁾

1) Белорусский государственный университет, Минск, Беларусь

E-mail: borzdov@bsu.by

2) ОАО «Интеграл» – управляющая компания холдинга «Интеграл», Минск, Беларусь

С помощью многочастичного метода Монте-Карло проведено моделирование тока в кремниевых фотодиодах с pin-структурой при обратном смещении и воздействии пикосекундных импульсов излучения с длиной волны 532 и 650 нм. Рассмотрены диодные структуры с различной толщиной нелегированной области.

Ключевые слова: метод Монте-Карло; фототок; кремниевый фотодиод.

ВВЕДЕНИЕ

Полупроводниковые фотодиоды широко применяются в качестве преобразователей оптического излучения в электрические сигналы. Кремниевые фотодиоды могут использоваться как детекторы излучения видимого и инфракрасного диапазонов и работать как в обычном токовом режиме, так и в режиме счета фотонов [1, 2].

Численное моделирование электрических характеристик фотодиодов является важной задачей в рамках разработки и проектирования таких элементов с целью прогнозирования их рабочих характеристик. Для моделирования электрических характеристик интегральных субмикронных фотодиодов и фототранзисторов одним из наиболее перспективных методов является самосогласованное моделирование на основе многочастичного метода Монте-Карло [3–6]. Метод Монте-Карло позволяет проводить моделирование процессов переноса носителей заряда в приборных полупроводниковых структурах с учетом всех значимых процессов рассеяния носителей заряда, а также процессов оптической генерации и лавинного умножения. В представленной работе проведено численное моделирование отклика тока на пикосекундные импульсы излучения видимого спектра в кремниевых фотодиодах с pin-структурой, работающих в режиме обратного смещения.

МОДЕЛЬ ФОТОДИОДА

Схематически структура моделируемого фотодиода с pin-структурой представлена на рис. 1. Излучение направлено нормально к поверхности

и охватывает область между электродами длиной $L_c = 0,25$ мкм. Толщина структуры вглубь подложки $W_s = 0,5$ мкм. Концентрации донорной и акцепторной примесей в n- и p-областях при расчетах полагались равными 10^{24} м^{-3} . Рассмотрены диоды с симметричной pin-структурой с $W_{d1} = W_{d2} = 50$ нм, 100 нм и 150 нм (толщины i-областей, соответственно, составляют 400 нм, 300 нм и 200 нм). Температура кристаллической решетки для моделирования задавалась равной 300 К.

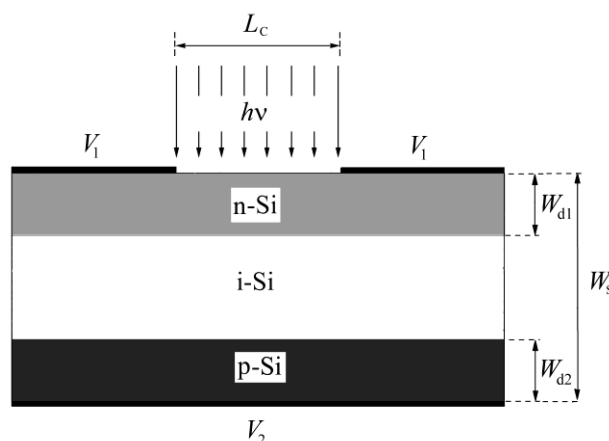


Рис. 1. Схематическое поперечное сечение кремниевого pin-фотодиода

Расчет фототока в рассмотренных диодных структурах проводился в рамках самосогласованного двумерного моделирования методом частиц, основы которого изложены в [7]. Интенсивность излучения составляла $5 \cdot 10^{10} \text{ Вт/м}^2$, длительность импульса — 1 пс. Излучение включалось в момент времени $t = 1$ пс. В качестве модельного приближения полагалось, что интенсивность импульсов постоянна во времени и равномерна по площади поверхности диодной структуры. Также полагалось, что коэффициент отражения на внутренних границах раздела рабочей области прибора (полупроводник–электрод подложки и полупроводник–внешняя среда) равен 1, при этом все внешнее излучение проникает внутрь рабочей области прибора (для падающего излучения нет отражения). Значения коэффициентов собственного поглощения в кремнии для рассмотренных длин волн, которые полагались независимыми от уровня легирования, взяты из [8].

Моделирование процессов переноса включало X- и L-долины для электронов, а также зоны тяжелых, легких и спин-отщепленную зону для дырок. При моделировании переноса носителей заряда учитывались процессы рассеяния электронов и дырок на акустических и оптических фононах, плазмонах и ионизированной примеси. В процедуру моделирования были включены процессы оптической генерации электронно-дырочных

пар при собственном поглощении излучения с учетом процессов отражения от границ раздела материалов. Процесс ударной ионизации электронами и дырками учитывался как отдельный механизм рассеяния с генерацией электронно-дырочных пар [9, 10].

РЕЗУЛЬТАТЫ РАСЧЕТОВ

На рис. 2 приведены зависимости отклика плотности тока от времени на пикосекундные импульсы излучения с длиной волны 532 нм (рис. 2а) и 650 нм (рис. 2б) в диоде с *pin*-структурой при обратном смещении $V_2 - V_1 = -10$ В.

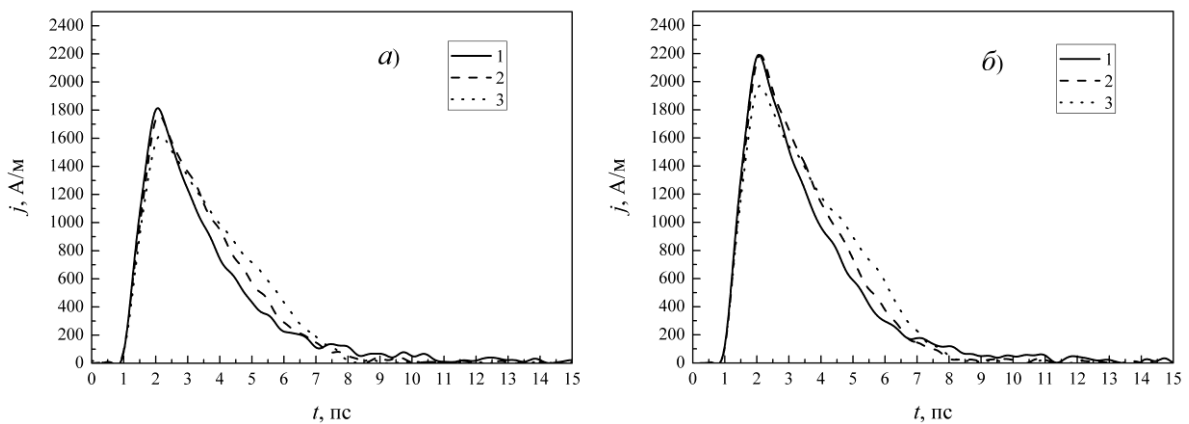


Рис. 2. Зависимости плотности тока от времени в диоде с *pin*-структурой при обратном смещении -10 В

На рис. 2 кривые 1 соответствуют диоду с *pin*-структурой при $W_{d1} = W_{d2} = 150$ нм (толщина *i*-области $W_i = 200$ нм), кривые 2 – $W_{d1} = W_{d2} = 100$ нм ($W_i = 300$ нм) и кривые 3 – $W_{d1} = W_{d2} = 50$ нм ($W_i = 400$ нм). Как видно из рисунков, для диодов с большей толщиной *i*-области наблюдается меньшее значение пикового значения тока и немного более медленное его затухание после окончания воздействия импульса излучения. Такое поведение можно объяснить тем, что с увеличением толщины *i*-области уменьшается напряженность электрического поля в ней, что, в свою очередь, приводит к уменьшению значения фототока. Также для рассмотренных условий отражения излучения (все падающее излучение поглощается в рабочей области прибора) величины фототоков получаются меньше при поглощении излучения с длиной волны 532 нм, нежели 650 нм. Связано это, предположительно, с тем, что при поглощении фотонов с длиной волны 532 нм средняя энергия сгенерированных электронно-дырочных пар в среднем оказывается выше, и в этом случае увеличивается интенсивность процессов рассеяния носителей заряда,

прежде всего на фонах, приводящих к междолинным переходам для электронов и межзонным для дырок. Последнее, в свою очередь, приводит к уменьшению подвижности носителей заряда.

ЗАКЛЮЧЕНИЕ

С помощью численного самосогласованного моделирования на основе многочастичного метода Монте-Карло проведено моделирование фототока в кремниевых фотодиодах с *pin*-структурой при воздействии пикосекундных импульсов излучения с длиной волны 532 и 650 нм и интенсивностью $5 \cdot 10^{10}$ Вт/м². Установлены закономерности динамики затухания тока в фотодиодах после окончания воздействия импульсов излучения.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Филачев А. М., Таубкин И. И., Трищенко М. А. Твердотельная фотоэлектроника. Фотодиоды // М.: Физматкнига, 2011. С. 448.
2. Modern methods of detecting single photons and their application in quantum communications / Koziy A. A., Losev A. V., Zavodilenko V. V. [et al.] // Quantum Electronics. 2021. Vol. 51, No. 8. P. 655–669. DOI: 10.1070/QEL17566
3. *Moglestue C.* Monte Carlo simulation of semiconductor devices // Wien: Springer, 2013. P. 334.
4. Full-band Monte Carlo simulations of photo excitation in silicon diode structures / Aboud S., Saraniti M., Goodnick S. [et al.] // Semicond. Sci. Technol. 2004. Vol. 19. P. S301–S303. DOI: 10.1088/0268-1242/19/4/101
5. *Yanikgonul S., Leong V., Ong J. R.* 2D Monte Carlo simulation of silicon waveguide-based single-photon avalanche diode for visible wavelengths // Optics Express. 2018. Vol. 26, No. 12. P. 15232–15246. DOI: 10.1364/OE.26.015232
6. *Borzdov A. V., Borzdov V. M., Vyurkov V. V.* Monte Carlo simulation of picosecond laser irradiation photoresponse of deep submicron SOI MOSFET // Proc. SPIE. 2022. Vol. 12157. P. 121570Y-1–121570Y-6. DOI: 10.1117/12.2624174
7. *Борздов В. М., Жевняк О. Г., Комаров Ф. Ф., Галенчик В. О.* Моделирование методом Монте-Карло приборных структур интегральной электроники // Минск: БГУ, 2007. С. 175.
8. *Wang H., Liu X., Zhang Z. M.* Absorption coefficients of crystalline silicon at wavelengths from 500 nm to 1000 nm // International Journal of Thermophysics. 2013. Vol. 34, P. 213–225. DOI: 10.1007/s10765-013-1414-2
9. Impact ionization model for full band Monte Carlo simulation / Kamakura Y., Mizuno H., Yamaji M. [et al.] // Journal of Applied Physics. 1994. Vol. 75, No. 7. P. 3500–3507. DOI: 10.1063/1.356112
10. A model of impact ionization due to the primary hole in silicon for a full band Monte Carlo simulation / Kunikiyo T., Takenaka M., Morifuji M. [et al.] // Journal of Applied Physics. 1996. Vol. 79, No. 10. P. 7718–7725. DOI: 10.1063/1.362375

МОДЕЛИРОВАНИЕ РЕЗОНАТОРА ДЛЯ ИЗМЕРЕНИЯ ВЛАЖНОСТИ ЗЕРНА

П.Д. Вольская, Г.Н. Дещенко

*Белорусский государственный университет, Минск, Беларусь
E-mail: polinavolskaya11@gmail.com*

В программной среде COMSOL Multiphysics создана модель кольцевого СВЧ резонатора. Получены АЧХ для различных значений диэлектрической проницаемости измеряемого материала. Изменение резонансных параметров можно использовать для измерения влажности зерна.

Ключевые слова: измерение влажности; резонансный метод; диэлектрическая проницаемость; амплитудно-частотная характеристика; COMSOL Multiphysics.

ВВЕДЕНИЕ

Мукомольные и зерносушильные предприятия нуждаются в технологичных методах контроля влажности зерновых культур. По сравнению с лабораторными измерениями, датчики влажности в потоке позволяют непрерывно отслеживать изменения параметров. Один из методов, используемых для таких измерений – резонансный. В литературе описана зависимость диэлектрических параметров зерна от влажности [2], однако их влияние на резонансные характеристики кольцевого резонатора не изучалось. Для исследования влияния влажности зерна на характеристики резонатора было проведено компьютерное моделирование датчика серии А315 [1]. Модель позволит уменьшить число практических экспериментов и упростит модернизации данных приборов.

МОДЕЛЬ РЕЗОНАТОРА

Сыпучий материал можно представить как смесь сухого вещества, воды и воздуха [4]. Как известно, диэлектрическая проницаемость воды $\epsilon \approx 81$, что значительно выше, чем сухого вещества ($\epsilon \approx 4$) и воздуха ($\epsilon \approx 1$), а значит её можно использовать для измерения содержания воды.

Конструкция резонатора представляет собой прямоугольный волновод с узкой щелью вдоль широкой стенки, свернутый в кольцо. При этом часть электромагнитного поля выходит и взаимодействует с исследуемым веществом [1]. При увеличении комплексной диэлектрической проницаемости вещества, ожидается уменьшение резонансной частоты и снижение амплитуды резонанса. Конструкция резонатора в разрезе представлена на рис. 1. Модель создана в программной среде COMSOL Multiphysics 6.1. Данные о комплексной диэлектрической проницаемости исследуемого материала взяты из [2] и соответствуют озимой пшенице.

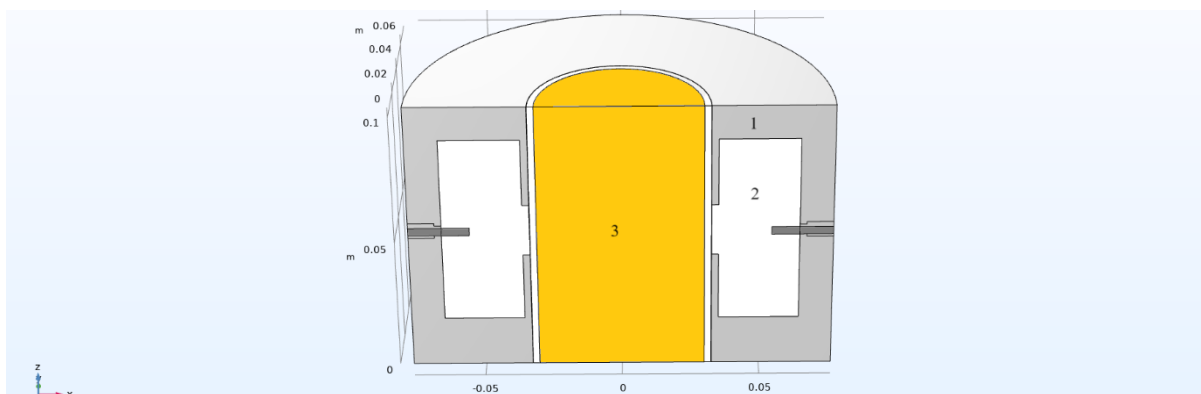


Рис. 1. Конструкция кольцевого резонатора в разрезе:
1 – металлическая граница резонатора, 2 – диэлектрический наполнитель (полиэтилен), 3 – исследуемый материал

РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ

Было проведено моделирование в частотной области для различных значений диэлектрической проницаемости исследуемого вещества, соответствующих воздуху, т.е. пустому резонатору, и пшенице влажности от 3% до 24%. Действительная часть диэлектрической постоянной изменялась от 2,2 до 3,6, а мнимая часть – от 0,11 до 0,41. Пример распределения напряженности электрического поля представлен на рис. 2. На резонансной частоте возникают стоячие волны, а напряженность поля резко возрастает.

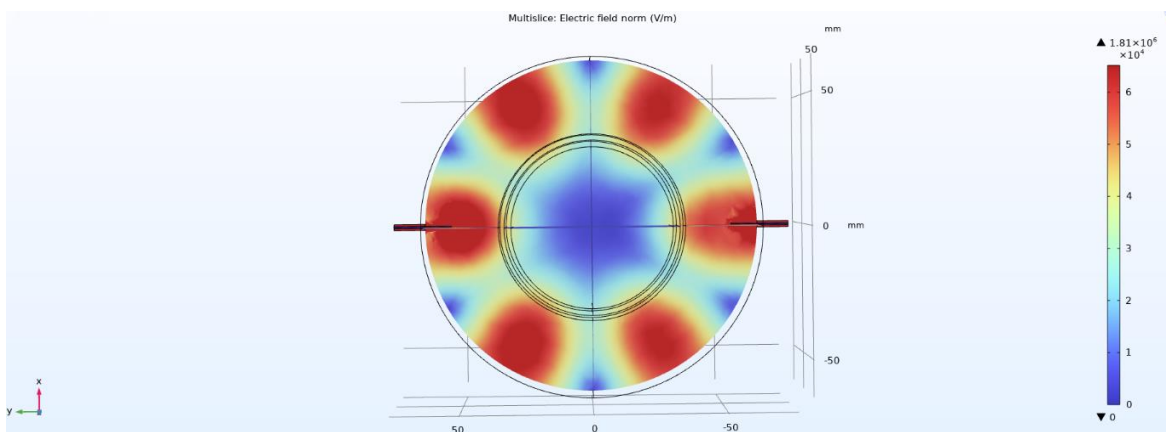


Рис. 2. Напряженность электрического поля в разрезе кольцевого резонатора. Частота 2,37 ГГц, диэлектрическая проницаемость исследуемого вещества $\epsilon = 1$ (воздух)

Для исследуемого вещества каждой влажности была получена АЧХ (рис. 3). С увеличением содержания воды частота уменьшается. Амплитуда уменьшается при низкой влажности, а на высокой начинает расти.

Из-за этого использование сдвига амплитуды для оценки влажности возможно только в ограниченном диапазоне, что находит отражение в алгоритмах оценки влажности данным методом при практической реализации датчика серии А315.

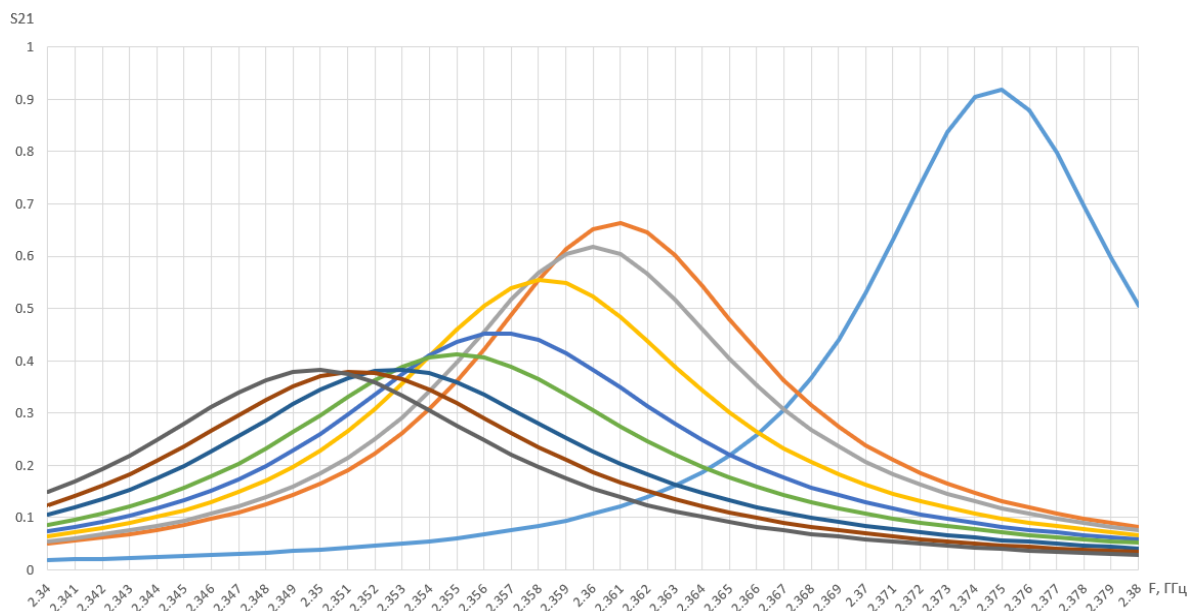


Рис. 3. АЧХ для незаполненного резонатора и для заполненного пшеницей влажности 3%, 6%, 9%, 12%, 15%, 18%, 21%, 24% справа налево соответственно

ЗАКЛЮЧЕНИЕ

В рамках исследования была построена компьютерная модель кольцевого резонатора, лежащего в основе датчика А315 и получены АЧХ, соответствующие зерну различной влажности. Результаты моделирования показали, что с увеличением влажности частота резонанса монотонно убывает, а амплитуда резонанса изменяется более сложным образом.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Беляниц А. Ч.* Микроволновый влагомер для зерна в потоке А315 для зерносушильных комплексов // Поволжье-Агро. 2020. №8 (127). С. 1-11.
2. ASAE D293.2 JUN1989 (R2005). Dielectric Properties of Grain and Seed // American Society of Agricultural and Biological Engineers (ASABE) Standards. 2006. С. 592-601.
3. COMSOL Multiphysics. RF Module User's Guide // COMSOL Documentation. 2023.
4. *Stuart O. Nelson.* Dielectric Properties and Density Relationships for Granular Materials // An ASAE/CSAE Meeting Presentation. 2004. С.2-4.

ОРИЕНТИРОВКА ВНУТРИ ЗДАНИЙ ПО УПРОЩЕННОМУ ПЛАНУ

Г.В. Гренкин¹, В.В. Карпов¹, А.Е. Чусова²

1 Владивостокский государственный университет, Владивосток

2 Дальневосточный федеральный университет, Владивосток

E-mail: Gleb.Grenkin@vvsu.ru; dev.karpov0@gmail.com; chusova.ae@dvfu.ru

В работе предлагается формат представления плана здания, упрощающий ввод плана и обеспечивающий достаточное количество сведений для ориентировки в здании. План здания представляется в виде прямолинейных участков движения (путей), ориентированных в четырех направлениях. Пути проходят мимо отдельных объектов (пунктов) и соединены опорными пунктами (стыками). Реализован вывод кратчайшего маршрута в здании между двумя пунктами.

Ключевые слова: indoor-навигация; ориентировка внутри помещений; кратчайший маршрут.

ВВЕДЕНИЕ

Навигация внутри зданий представляет важную задачу, решение которой позволит сократить время в поисках нужного места. Отличие indoor-навигации от уличной навигации состоит в необходимости для пользователя читать схемы. В настоящей работе поставлена задача построения модели представления данных о зданиях, которая использует минимальную информацию о плане и при этом допускает построение кратчайших маршрутов между заданными местами. Отсутствие графического представления схемы здания компенсируется внятным диалогом с пользователем и вставками фотографий опорных пунктов.

Решение задачи навигации можно разделить на несколько этапов:

- определение местоположения человека;
- построение оптимального пути до точки назначения;
- генерация описания маршрута.

Для решения задачи автоматического определения местоположения человека в здании могут использоваться штрих-коды [1] или локационные маячки [2]. Задача построения текстового описания маршрута ранее рассматривалась в работе [3], в которой использовались данные Google Maps. В [4] для обеспечения ориентировки в помещении построена трехмерная модель здания со сложной структурой.

Работа [5] решает комплексную проблему создания мобильного навигационного сервиса. В [6] отмечают, что для навигации требуется сочетание геометрических, логических и семантических моделей, и предлагают модель навигационного пространства, разделенного на ячейки, с помо-

щью которой строится логическая модель здания. Авторы [7, 8] предложили расширение формата OpenStreetMap на внутренние планы зданий, для чего в формат были добавлены теги, отвечающие за структуру помещений. В статье [9] разработана модель представления данных о маршрутах внутри и вне помещений на основе этого формата. Авторы обращают внимание на проблему генерации текстовых маршрутов, решаемую посредством внедрения предметно-ориентированного языка [10]. Отметим также работу [11], в которой для упрощенного представления плана использовался формат OpenStreetMap. Формат OpenStreetMap обладает определенной избыточностью, которая уменьшает его читабельность и создает препятствия для разработчиков карт. В то время как данный формат будет эффективен для крупного здания, для остальных он может оказаться чересчур громоздким.

Целью нашей работы является разработка эффективного формата представления планов зданий для ускорения ввода плана и использования.

ФОРМАТ ПЛАНА ЗДАНИЯ

Для представления плана в упрощенном формате выделяются следующие сущности:

1. Путь (path) – прямолинейный участок маршрута, доступный для перемещения пешехода, обладающий параметрами направления и начального пункта. Путь может быть корневым или производным. Первый начинается с корневого пункта, второй стартует со стыка и определённым образом соотносится с родительским путём. Завершается путь стыком или тупиковым пунктом.

2. Пункт (point) – место, мимо которого проходит путь. Он обладает параметрами: идентификатор, название и стенка, со стороны которой он находится. Корневой пункт относится непосредственно к этажу, где он расположен. В качестве корневого пункта может быть выход, лестница, лифт или эскалатор.

3. Стык (joint) – опорный пункт, который служит поворотом или развилкой. Стык имеет параметры: название, текст в зависимости от направления движения и фото в зависимости от направления движения.

Для представления указанных сущностей и отношений между ними используется формат XML, содержащий теги Floor, в которые вложены теги корневых пунктов Point и теги корневых путей Path. Каждый тег Path содержит последовательность вложенных тегов Point, которая заканчивается тегом Joint, после которого могут следовать вложенные теги Path, соответствующие производным путям.

Для обозначения вертикальных перемещений вводятся теги Stairs (лестница) и Elevator (лифт). На каждом этаже лестницы и лифты имеют

одинаковые идентификаторы, что позволяет соединить объекты на разных этажах ребрами при обработке структуры данных.

На основе введенного формата строится структура данных «План». Эта структура данных дополняется ребрами, соединяющими каждую пару соседних пунктов, принадлежащих общему пути.

Для построения кратчайшего маршрута в не взвешенном графе применяется волновой алгоритм. Затем для построенного маршрута генерируется текстовое описание. Для этого выделяются фрагменты маршрута, относящиеся к одному пути. Эти фрагменты выводятся в виде отдельного блока: «Поверните <в направлении>, пройдите до <пункта> и <стыка> [слева: <пункты слева>; справа: <пункты справа>]».

Ключевые отличия предлагаемого формата от OpenStreetMap:

- 1) минималистичность: формат содержит только основную информацию, не требуя ввода геометрической модели здания;
- 2) формат ориентирован на пути, по которым движется пешеход, в то время как в плане OSM такая информация опциональна;
- 3) специально выделяются ключевые пункты (стыки), содержащие информацию о поворотах, специальный текст и фотографии.

ВЫВОДЫ И ПЕРСПЕКТИВЫ

Возможно преобразование формата OpenStreetMap (OSM) в предлагаемый формат. Формат OSM включает следующие элементы: точки (node), линии (way), отношения (relation), а также метки (tags). Если выделить пути для перемещения пешеходов, указать направления путей и обозначить объекты, расположенные вдоль них, то получим структуру данных, соответствующую нашему формату. Формат OSM допускает ввод пешеходных путей посредством установки атрибута corridor у соответствующих линий. Ввод таких линий должен упростить преобразование формата OSM в представленный формат.

На основе разработанного формата может быть создан веб-сервис для совместного редактирования планов зданий с последующим использованием для навигации. Предлагается осуществлять изменение схемы напрямую в формате XML, используя инструменты, упрощающие этот процесс. Результат может отображаться на отдельной панели в виде схемы с указанием путей, пунктов и стыков. На начальном этапе внедрения этой технологии набор XML-файлов с планами может размещаться в открытом репозитории сервиса GitHub.

Предложенный нами формат будет предпочтительнее в тех приложениях, где геометрическая точность не столь существенна. В дальнейшем может быть востребовано создание редактора данного формата с функ-

цией схематичного отображения плана здания с целью обнаружения ошибок в плане. Программная реализация описанной системы доступна в репозитории по адресу <https://github.com/lapkin25/Nav-any>.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Абдрахманова А.М., Намиот Д.Е. Использование двумерных штрихкодов для создания системы позиционирования и навигации в помещении // Прикладная информатика. 2013. № 1 (43). С. 31-39.
2. Колышкина Н.С., Борсук Н.А. Анализ решения вопросов indoor-навигации // Южно-Сибирский научный вестник. 2021. № 2 (36). С. 98-103.
3. Пестун М. В. Когнитивная навигация и алгоритм построения текстового описания маршрута в удобном для человека виде // Программные продукты и системы. 2015. № 1 (109). С. 28-33.
4. Охотниченко А.В., Кухта Ю.Б. Проектирование системы для навигации внутри здания со сложной иерархической структурой // Программные системы и вычислительные методы. 2021. № 4. С. 46-57.
5. Дусакаева С.Т., Савинов В.В. Мобильное приложение для поиска оптимального маршрута в университетском городке // Онтология проектирования. 2023. Т. 13. – № 3 (49). С. 455-464.
6. Liu L., Zlatanova S. Generating navigation models from existing building data // The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2013. V. 40. P. 19-25.
7. Goetz M., Zipf A. Extending OpenStreetMap to indoor environments: bringing volunteered geographic information to the next level // Urban and Regional Data Management: UDMS Annual. 2011. V. 2011. P. 47-58.
8. Goetz M., Zipf A. Indoor route planning with volunteered geographic information on a (mobile) web-based platform // Progress in Location-Based Services. 2013. P. 211-231.
9. Wang Z., Niu L. A data model for using OpenStreetMap to integrate indoor and outdoor route planning // Sensors. 2018. V. 18. № 7. P. 2100.
10. da Silva Andrade E., Vale S.B. Indoor-DSL: A model driven approach to modeling indoor representation models for a navigation infrastructure // International Journal of Computer Applications. 2020. V. 975. P. 8887.
11. A new spatial database framework for pedestrian indoor navigation based on the OpenStreetMap tag information / Dumedah G. [et al.] // Transactions in GIS. 2022. Vol. 26, № 7. P. 3090–3108.

МОДЕЛИРОВАНИЕ ВЕЛИЧИНЫ ПАРАЗИТНОГО ТУННЕЛЬНОГО ТОКА В ЭЛЕМЕНТАХ ФЛЕШ-ПАМЯТИ

Жевняк О. Г., Борздов В. М., Борздов А. В.

Белорусский государственный университет, Минск, Беларусь

Email: Zhevnyakol@tut.by

В настоящей работе с помощью численного моделирования методом Монте-Карло электронного переноса в короткоканальных МОП-транзисторах с плавающим затвором, лежащих в основе работы современных элементов флеш-памяти, рассчитаны зависимости относительной величины паразитного туннельного тока от координаты вдоль проводящего канала транзистора. Рассмотрено влияние на них изменения формы потенциального барьера вдоль канала, а также энергии и дрейфовой скорости электронов. Показано, что эти зависимости имеют слабо выраженный U-образный вид.

Ключевые слова: флеш-память; туннельный ток; метод Монте-Карло.

Одной из основных тенденций современных флеш-технологий является уменьшение размеров элементов флеш-памяти [1]. Но при этом недопустимо повышение паразитных туннельных токов в них, приводящих к искажению хранящейся в этих элементах информации (см., например, [2]). Целью настоящего исследования явилось моделирование методом Монте-Карло пространственного распределения относительной величины паразитных туннельных токов в короткоканальных МОП-транзисторах с плавающим затвором, являющихся базовым элементом флеш-памяти, вдоль их проводящего канала. Алгоритмы такого моделирования описаны в работе [3], а процедура расчета туннельного тока из проводящего канала на плавающий затвор рассмотрена в работе [4].

На рис. 1 приведена конструктивная схема моделируемого элемента флеш-памяти. В настоящей работе приведены результаты моделирования для прибора со следующими конструктивно-технологическими параметрами: длина канала $L_{ch} = 0,2$ мкм, толщина подзатворного окисла $d_{окс} = 7$ нм, толщина туннельного окисла $d_{тун} = 2$ нм, толщина плавающего затвора $d_{плав} = 2$ нм, глубины залегания истоковой и стоковой областей в подложку $d_j = 50$ нм, концентрация донорной примеси в областях истока и стока $N_D = 10^{26}$ м⁻³, напряжения на стоке и затворе $V_D = 2$ В, $V_G = 2$ В.

Паразитный туннельный ток образуют электроны, находящиеся в проводящем канале у границы раздела Si/SiO₂ и туннелирующие на плавающий затвор через туннельный оксид. Плотность этого тока численно может быть рассчитана согласно формуле

$$j = eN_n D v dr ,$$

где e – заряд электрона, N_n и v_{dr} – концентрация и дрейфовая скорость электронов, находящихся у границы раздела Si/SiO₂, D – коэффициент туннелирования, определяемый величиной энергии электронов E , высотой барьера H на границе раздела Si/SiO₂ и разностью уровней дна зон проводимости в канале и на плавающем затворе ΔU .

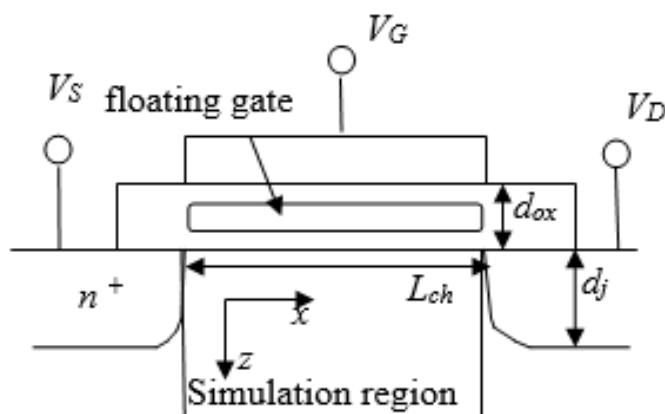


Рис. 1. Конструктивная схема рассматриваемого элемента флеш-памяти и область моделирования

Величины N_n , v_{dr} и E вычисляются с помощью стандартных процедур метода Монте-Карло. Величина ΔU прямо пропорциональна величине электрического потенциала в проводящем канале и определяется с помощью самосогласованного решения уравнения Пуассона. На рис. 2 приведены три диаграммы, поясняющие характер изменения этой величины вдоль канала и показано, как она определяет форму потенциального барьера между плавающим затвором и каналом, который и регулирует процесс формирования паразитного туннельного тока.

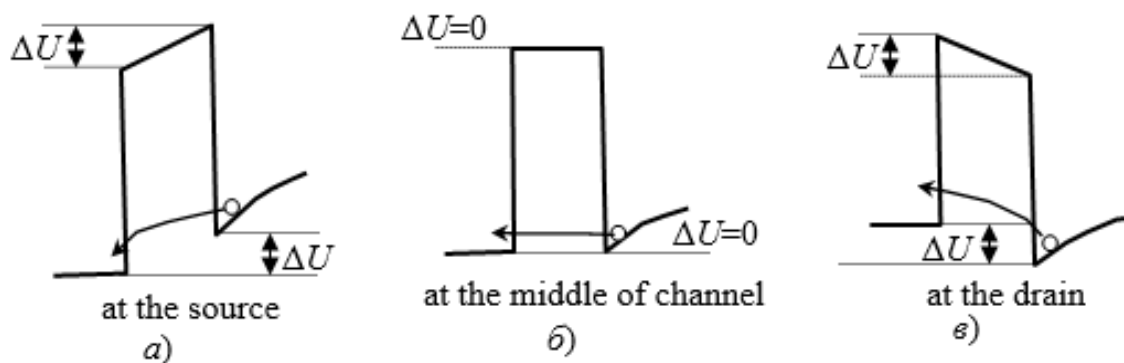


Рис. 2. Энергетическая диаграмма изменения вида барьера на границе Si/SiO₂ вдоль проводящего канала:

а - у истока, б - в середине канала, в - у стока

Высота барьера H для рассматриваемого элемента флеш-памяти равнялась 3,2 эВ. Значения величин N_n , v_{dr} , E и ΔU существенно изменяются вдоль канала, при этом величина ΔU , как легко видеть из рис. 2, вообще меняет свой знак на противоположный. Поведения этих величин и определяют закономерности пространственного распределения величины паразитного туннельного тока в элементах флеш-памяти.

Нами рассчитаны зависимости вдоль канала относительной величины паразитного туннельного тока для МОП-транзисторов с плавающим затвором для приборов с разной длиной проводящего канала, глубиной залегания истоковой и стоковой областей, концентрацией акцепторной области. Для примера на рис. 3, б приведена полученная зависимость $J_{\text{tun}}/J_{\text{drain}}$ для транзистора с $L_{ch} = 0,2$ мкм, $d_j = 50$ нм и концентрацией $N_A = 10^{24} \text{ м}^{-3}$. Для сравнения на этом же рисунке (рис. 3, а) приведена зависимость изменения величины коэффициента D , рассчитанная для среднестатистических электронов, имеющих среднюю энергию и находящихся у границы раздела Si/SiO₂ в соответствующем сечении канала.

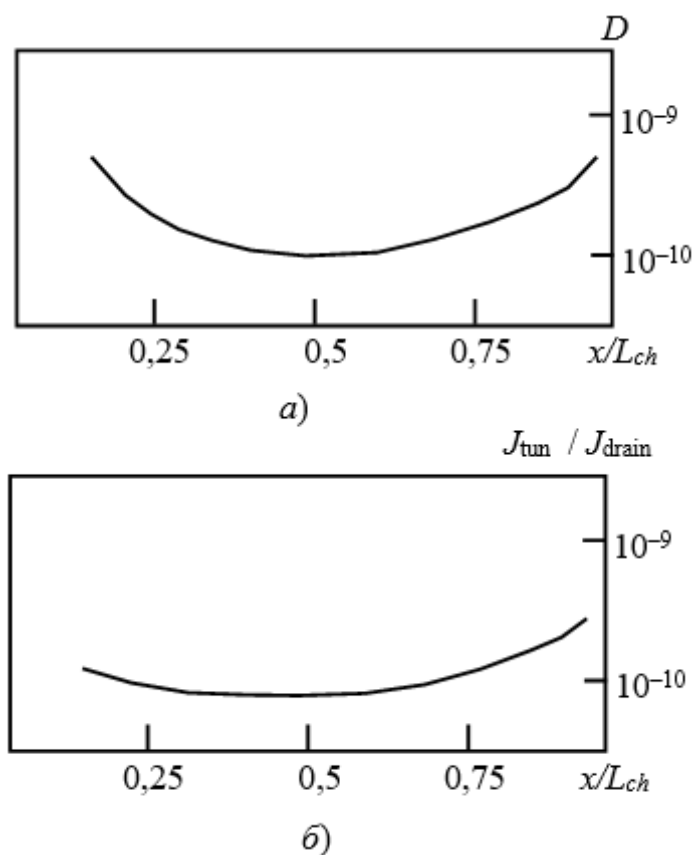


Рис. 3. Изменения вдоль канала моделируемого элемента флеш-памяти значений коэффициента туннелирования (а) и относительной величины туннельного тока на плавающий затвор (б)

Между кривыми прослеживается очевидная корреляция. Характерная U-образная зависимость коэффициента D обусловлена изменением знака ΔU с положительного на отрицательный при изменении потенциала в канале от истока к стоку (приводит к уменьшению D) и ростом энергии электронов E вдоль канала (приводит к увеличению D) – на начальном участке преобладает первое явление, на конечном второе. Для относительной величины паразитного туннельного тока U-образная зависимость выражена несколько слабее. Это обусловлено тем, что на данную величину помимо коэффициента D влияние оказывает ряд других параметров и, прежде всего, величина дрейфовой скорости, которая в начале канала очень мала.

Таким образом, в настоящей работе с помощью численного моделирования электронного переноса методом Монте-Карло рассчитаны зависимости вдоль канала относительной величины паразитного туннельного тока. Установлено, что эти зависимости имеют слабо выраженный U-образный вид, обусловленный особенностями паразитного туннелирования электронов в рассматриваемых элементах, и изменениями вдоль проводящего канала транзистора величин концентрации электронов, их энергии и дрейфовой скорости.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Hu C. (ed.) Nonvolatile semiconductor memories: Technologies, design, and applications.* Piscataway, NJ: IEEE Press, 1991. P. 167.
2. *De Salvo B. Silicon Non-Volatile Memories: paths of innovation.* London: Wiley-ISTE Ltd, 2009. P. 256.
3. *Борздов В.М., Жевняк О.Г., Комаров Ф.Ф., Галенчик В.О.* Моделирование методом Монте-Карло приборных структур интегральной электроники. Минск: БГУ; 2007. С. 175.
4. *Жевняк О. Г.* Моделирование паразитных туннельных токов в элементах флеш-памяти на основе короткоканальных кремниевых МОП-транзисторов // Eurasian union of Scientists (ECSU). 2020. № 76, Ч. 2. С. 26–28.

КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ ТЕПЛООБМЕННИКА ПЕЧИ ПИРОЛИЗА ДЛЯ ОПТИМИЗАЦИИ ПРОЦЕССА ПИРОЛИЗА УГЛЕВОДОРОДОВ

М.Е. Захаров, Т.В. Ганджа

*Томский государственный университет систем управления и радиоэлектроники,
Томск, Российская Федерация
E-mail: epsn97@mail.ru*

Данная научная статья посвящена использованию компьютерного моделирования для оптимизации процесса пиролиза углеводородного сырья через анализ теплообменника печи пиролиза. Рассматриваются современные подходы к моделированию теплообмена, его влияние на эффективность процесса пиролиза и возможности оптимизации теплообменника для улучшения производственных показателей.

Ключевые слова: компьютерное моделирование; оптимизация; пиролиз углеводородов.

ВВЕДЕНИЕ

Полимерная продукция в мировой экономике занимает весомое место, она является востребованной во многих областях жизнедеятельности современного общества – машиностроение, автомобилестроение, сельское хозяйство, космическая промышленность, военная промышленность, текстильная промышленность и другие. Благодаря широкому спектру использования спрос на полимеры ежегодно растёт.

Особую роль для получения ценных химических продуктов занимает процесс пиролиза углеводородов. Пиролиз включает в себя процессы высокотемпературного нагревания смеси веществ, реакции расщепления, конденсации, полимеризации, протекающих при низком давлении. Это сложный химико-технологический процесс, состоящий из совокупности химических реакций.

Одним из ключевых элементов этого сложного процесса является теплообменник печи пиролиза, который влияет на температурный режим и эффективность процесса. В данной работе исследуется применение компьютерного моделирования для оптимизации теплообменника и улучшения процесса пиролиза углеводородов.

Компьютерное моделирование играет важную роль в оптимизации технологического процесса, обеспечивая повышение эффективности производства, снижение затрат и улучшение качества продукции.

ТЕХНОЛОГИЧЕСКОЕ ОПИСАНИЕ ПЕЧИ ПИРОЛИЗА УГЛЕВОДОРОДОВ

Перед тем как в распределительный коллектор подать углеводородное сырьё (бензин), его необходимо нагреть. Нагрев происходит путём подачи сырья в межтрубное пространство теплообменников, где оно нагревается до температуры около $30^{\circ}\text{C} \div 100^{\circ}\text{C}$, после чего поступает в печь пиролиза. В камере конвекции печи пиролиза, которая состоит из шести секций, происходит ряд химико-технологических процессов: испарение сырья, подогрев смеси сырья и водяного пара, а также перегрев пара высокого давления. Прямогонный бензин подаётся четырьмя потоками в первую секцию камеры конвекции печи, где происходит нагрев углеводородного сырья. Расход сырья на каждый поток настраивается регулируемыми клапанами и зависит от температуры смеси перед закалочными-испарительными аппаратами. На данном этапе происходит реакция превращения олефинов в продукты полимеризации.

Для выхода на оптимальные условия протекания реакции в технологическом процессе предусмотрена подача водяного пара для разбавления сырья, что позволяет уменьшить парциальное давление.

После разбавления сырья, полученная паро-сырьевая смесь (водяной пар и бензин) с температурой около $100^{\circ}\text{C} \div 300^{\circ}\text{C}$, последовательно поступает в пятую и в шестую секции конвекционной камеры. На этих стадиях смесь нагревается до температуры $300^{\circ}\text{C} \div 600^{\circ}\text{C}$, после чего направляется в камеру радиации печи.

ВЛИЯНИЕ КОНСТРУКТИВА ПЕЧИ НА ПРОТЕКАНИЕ ПРОЦЕССА ПИРОЛИЗА

Одна из важных характеристик, оказывающей влияние на процесс пиролиза – диаметр труб печи, который связан с температурой внешней поверхности трубы прямо пропорциональной зависимостью. В результате отложения слоя кокса температуры внешней и внутренней поверхности труб возрастают и скорость отложения кокса увеличивается [1].

Отложение кокса в процессе пиролиза углеводородного сырья необходимо минимизировать. Для решения данной задачи необходимо оптимально подобрать длину и диаметр труб на печах пиролиза, путём компьютерного моделирования теплообменника печи.

МЕТОДИКА МОДЕЛИРОВАНИЯ И ОПИСАНИЕ СРЕДЫ МОДЕЛИРОВАНИЯ

Теплообменник – это аппарат ХТС, в котором осуществляется теплообмен между двумя теплоносителями – многокомпонентными вещественными потоками, находящимися в газовой или жидкой фазе и имеющими различную температуру. В нём осуществляется перенос теплоты (теплоперенос) от более

нагретого теплоносителя, называемого далее горячим, к менее нагретому – холодному, через разделяющую их стенку [2].

Компьютерная модель теплообменника печи пиролиза учитывает теплопередачу, тепловые потери и гидродинамические параметры. Моделирование теплообменника реализовано в среде математического моделирования MAPS, разработанной в Томском государственном университете систем управления и радиоэлектроники (ТУСУР). MAPS была разработана на основе развития метода компонентных цепей, организующего параллельно взаимодействующие уровни моделирования – объектный, логический и визуальный. Метод компонентных цепей [3] является универсальным методом компьютерного моделирования физически неоднородных объектов в статическом и динамическом (во временной и частотной областях) режимах. Среда многоуровневого моделирования MAPS разработана на языке Visual C++ с применением технологии объектно-ориентированного программирования (ООП) и набора классов библиотеки Microsoft Foundational Classes (MFC). MAPS имеет открытую библиотеку моделей компонентов, в которой реализованы модели основных элементов и аппаратов химической промышленности, а также модели компонентов, обеспечивающие проведение различных видов вычислительного эксперимента, визуализации его результатов и управления его ходом и параметрами анализируемых моделей [2]. Схема модели теплообменника представлена на рис. 1.

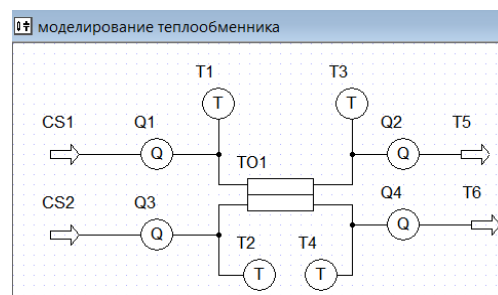


Рис. 1. Схема модели теплообменника

В теплообменник TO1 подаётся два потока. CS1 – холодный поток (прямогонный бензин), CS 2 – горячий поток (паро-сырьевая смесь). Для контроля процесса на схеме установлены следующие элементы: Q – датчики теплового потока, T1-T4 – датчики температуры, T5,6 – терминаторы.

По результатам моделирования теплообменника (рис. 2), подтверждается прямая зависимость длины трубы и температуры протекания процесса. Под показателями (ДО) принимается длина трубы 11 м, которая считается распространённой на нефтеперерабатывающих заводах. Под показателями (ПОСЛЕ) предлагается длина трубы 9 м, отобранная в результате моделирования. По выходным параметрам моделирования видно, что более оптимальной длиной для уменьшения коксообразования и уменьшения температуры

протекания процесса, является длина трубы 9 м. Подобные преобразования позволяют улучшить качество выпускаемой продукции и увеличить срок службы печей пиролиза углеводородного сырья.

Q горячий ДО	T горячий ДО
5.920e+009	253.00
Q холодный ДО	T холодный ДО
4.850e+009	202.00
Q горячий ПОСЛЕ	T горячий ПОСЛЕ
5.715e+009	226.03
Q холодный ПОСЛЕ	T холодный ПОСЛЕ
5.010e+009	207.56

Рис. 2. Результаты моделирования

ЗАКЛЮЧЕНИЕ

Моделирование проводилось с учетом изменений внешних условий работы теплообменника. Результаты моделирования возможно применять для оптимизации конструкции теплообменника и улучшения процесса пиролиза углеводородного сырья.

Анализ результатов моделирования теплообменника показал, что оптимизация его конструкции может значительно повысить эффективность процесса пиролиза углеводородов. Изменения в геометрии теплообменника позволяют улучшить теплообмен и обеспечить оптимальные условия для протекания процесса.

Компьютерное моделирование теплообменника печи пиролиза является эффективным инструментом для оптимизации процесса пиролиза углеводородов. Разработка и применение компьютерных моделей позволяет улучшить теплообмен и повысить эффективность производства ценных химических продуктов. Дальнейшие исследования в данной области могут способствовать развитию новых технологий и методов оптимизации процесса пиролиза.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Пиролиз углеводородного сырья. / Мухина Т.Н. [и др.] // М.: Химия, 1987.
2. Дмитриев В.М., Ганджа Т.В. Среда многоуровневого компьютерного моделирования химико-технологических систем. // Томск: Издательство Том. унта, 2017. С. 332.
3. В.М. Дмитриев, Л.А. Арайс, А.В. Шутенков. Автоматизация моделирования промышленных роботов // М.: Машиностроение, 1995. С. 304.

ИДЕНТИФИКАЦИЯ СТОХАСТИЧЕСКОЙ МОДЕЛИ ТЕПЛОПОТРЕБЛЕНИЯ В МНОГОЭТАЖНЫХ ЖИЛЫХ ЗДАНИЯХ ГОРОДА

О.В. Стукач, Е.А. Карапиш

*Национальный исследовательский университет "Высшая школа экономики", г.
Москва, Россия
E-mail: karapishkat@gmail.com*

Рассматривается построение модели здания в виде линейной дискретной стохастической системы с неизвестными параметрами. Для поиска параметров представлен алгоритм на основе метода наименьших квадратов с возможностью контроля точности. Примеры работы алгоритма приведены на основе реальных данных теплоснабжения зданий города Томска и внешней среднесуточной температуры. Показана возможность применения модели здания для разных периодов времени, а также пригодность модели для разных классов жилых зданий.

Ключевые слова: идентификация системы; потребление тепла; поиск параметров; жилое здание; энергоэффективность.

ВВЕДЕНИЕ

Вопрос эффективного потребления тепла жилыми зданиями поднимается в научных работах с целью оптимизации расхода энергии [1, 2]. В ходе исследований по анализу данных теплоснабжения зданий было выявлено, что различные характеристики домов, например, материал стен и год постройки, в меньшей степени влияют на эффективность потребления тепла зданиями, в отличие от возможности регулировать его подачу [3].

В качестве объекта управления в работе создается модель в виде линейной дискретной стохастической системы на основе данных потребления тепла, взятых с общедомовых приборов учёта в городе Томске и среднесуточной температуры по данным гидрометцентра [4].

Для поиска параметров модели в качестве базового алгоритма использован метод наименьших квадратов с возможностью задания размерности системы. Однако метод не учитывает возможность оценки точности и ограничения на используемые данные [5], что при построении регулирования не может гарантировать оптимальный расход ресурсов. Поэтому в базовый алгоритм внесены изменения с добавлением критерия останова для учета наблюдений на основе среднеквадратичной точности и добавлен выбор размерности модели на основе минимизации средней ошибки регулирования.

АЛГОРИТМ ПОИСКА ПАМЕТРОВ МОДЕЛИ ЗДАНИЯ

Рассматриваемая модель здания имеет n уравнений, каждый из которых описывает потребление тепла за один день, учитывая предыдущие данные. Далее x_{k-1} – количество тепла за предыдущие k дней, u_{k-1} – среднее значение температуры, ε_k – ошибка регулирования.

$$x_k = a_1(k)x_{k-1} + \dots + a_p(k)x_{k-q} + b_1(k)u_{k-1} + \dots + b_q(k)u_{k-q} + \varepsilon_k, k = 1, 2, \dots \quad (1)$$

Построение базового алгоритма идентификации основано на применении к системе дифференциальных уравнений метода наименьших квадратов в матричном виде с модификациями, проведенными для конкретной модели, описанной выше.

Вектор параметров модели определяется по формуле

$$\hat{\theta}_n = (B^T B)^{-1} (B^T X), \quad (2)$$

где B – матрица коэффициентов при искомым параметрах, X – вектор реальных значений теплопотребления.

Полученный алгоритм в реализации требует ручного задания количества уравнений в системе n и количества дней k . Такой подход не дает однозначного описания процесса потребления тепла одним зданием.

Измененный алгоритм предлагает автоматическое определение количества берущихся в расчет дней за счет указания желаемой точности и максимальных значений n и k . Значение h – величина определяющая точность, ее обратное значение $1/h$ является ограничением для момента останова $\tau(h)$ на основе информационной матрицы $(B^T B)$

$$\tau(h) = \inf \{n \geq n_0 : ||(B^T B)^{-2}||^{1/2} \leq h^{-1}\}. \quad (3)$$

Для полученного значения $n = \tau(h)$ определяется значение k . Для каждого k составляется модель, описанная выше, и определяются значения неизвестных параметров. Значение, при котором средняя по модулю ошибка регулирования является минимальной, выбирается как оптимальное.

ПРИМЕНЕНИЕ АЛГОРИТМА

Для отдельного дома с УСПД 13 (устройство сбора и передачи данных) при применении базового алгоритма можно исследовать построенные модели при задании различных значений n и k для заданной точности.

На рис. 1а и б показаны графики, для построения которых бралось слишком малое и слишком большое количество данных теплоучета соответственно. На рис. 1в показан пример с автоматическим выбором оптимальных значений n и k при заданных максимальных $n_{\max}$, $k_{\max}$.

Видно, что неверно выбранные значения n , k приводят к потере точности. Наоборот, усовершенствованный алгоритм достаточно точно прогнозирует теплотребление на основе имеющихся данных, а также убивает необходимость ручного подбора параметров системы.

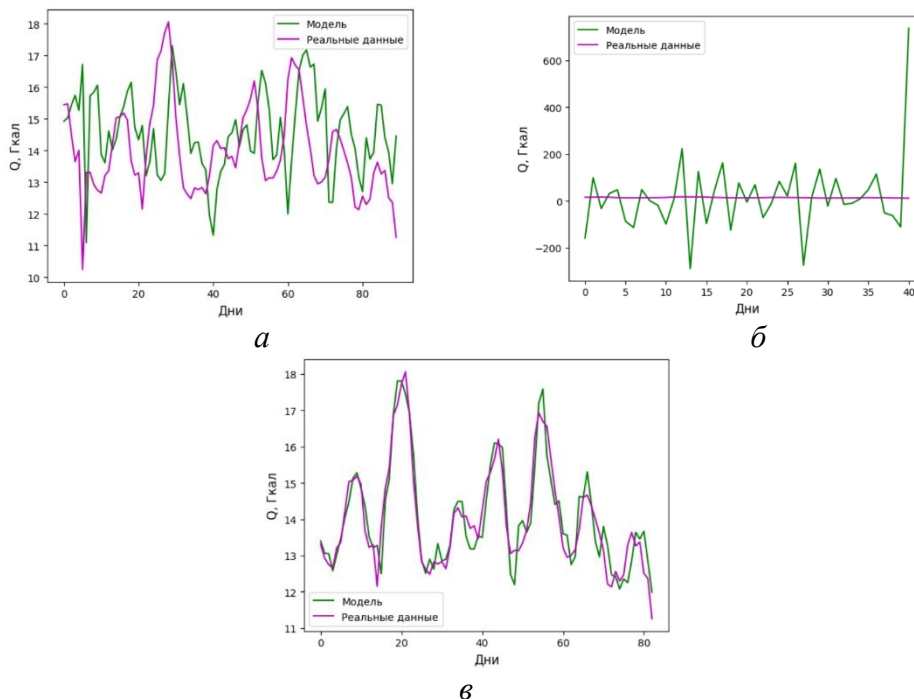


Рис. 1. Реальные данные теплотребления здания с УСПД 13 и графики для модели с различными n , k : $n = 4, k = 1$ – (а); $n = 90, k = 50$ – (б); $n_{\max} = 90, k_{\max} = 50, n = 50, k = 8$ – (в)

Подобранную модель можно также распространить на другой интервал времени для выбранного здания. На рис. 2, в отличие от предыдущего, берется следующий отопительный период.

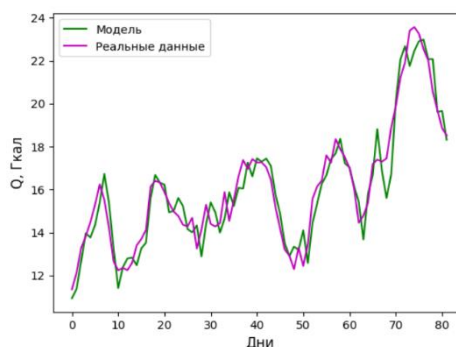


Рис. 2. Реальные данные теплотребления здания с УСПД 13 и модельные результаты на следующий отопительный период

Но подобная модель не является универсальной для любого здания. Например, при попытке описать, вышеприведенной моделью, процесс

теплопотребления для здания с УСПД 45 получается приближенное повторение поведения системы (рис. 3а), в отличие от использования модели, полученной на основе реальных данных этого здания (рис. 3б).

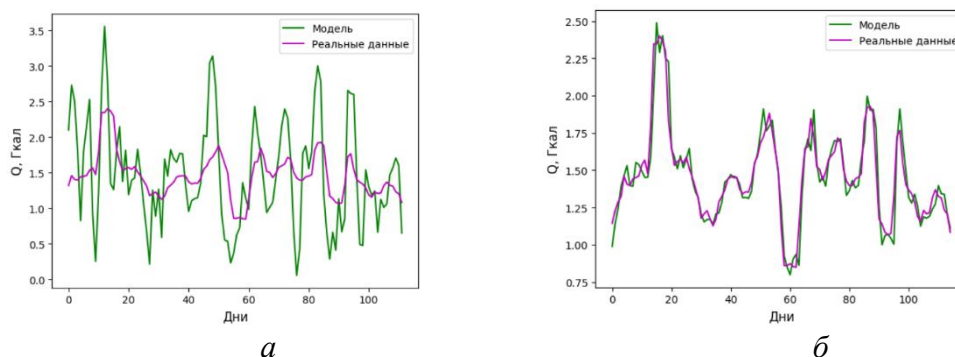


Рис. 3. Реальные данные теплопотребления здания с УСПД 45 и модель для здания: с УСПД 13 – (а), с УСПД 45 – (б)

ЗАКЛЮЧЕНИЕ

Представленный алгоритм построения модели здания и поиска параметров на основе реальных данных теплопотребления позволяет достаточно точно описывать изменение количества потребляемой энергии. Результаты работы могут быть использованы для построения модели регулятора, где в качестве объекта регулирования может быть здание, представленное подобной моделью.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Агаев Н. Б. Использование нечеткой прогностической модели в оперативно-диспетчерском анализе работы систем теплоснабжения // Теплоэнергетика. 2020. № 9. С. 107–112.
2. Хафизов Р. Н. Моделирование функционирования системы централизованного теплоснабжения // Вестник Самарского государственного технического университета. Серия: Технические науки. 2007. №2 (20).
3. Стукач О. В., Зорин П. А. Дисперсионный анализ данных коммерческого учёта тепловой энергии в жилом фонде города Томска // Наука. Технологии. Инновации: Сборник научных трудов. В 9-и частях, Новосибирск, 2019. С. 72-75.
4. Stukach O., Zorin P. Long-Term Data from the Heat Meters in Residential Buildings Depending on the Outside Temperature and Characteristics of Buildings // IEEE Data-port. 2021. DOI: 10.21227/cw53-rr81.
5. Кашковский Д.В., Конев В.В. Последовательная идентификация линейной динамической системы со случайными параметрами // Автоматика и телемеханика. – 2008. Вып. 8. С. 82–95.

ПРОЕКТИРОВАНИЕ КОРПОРАТИВНОЙ ИНФОРМАЦИОННОЙ СЕТИ КОМПАНИИ

О.Ф. Ковалёв, Т.Ф. Шмигирёв

БГУ, Минск, Беларусь
E-mail: OKovaliov@gmail.com

Компаниям, имеющим несколько филиалов, требуется организовать корпоративную информационную сеть для общей работы с базами данных и документами, позволяющую сотрудникам общаться и обмениваться данными, а также обеспечить безопасную передачу этих данных и защиту от проникновения извне. В статье представлено эффективное проектирование и тестирование корпоративной сети предприятия с учётом требований к безопасности, скорости сходимости и масштабируемости.

Ключевые слова: корпоративная информационная сеть; маршрутизатор; динамическая маршрутизация; OSPF; BDP; межсетевой экран; ContainerLab.

ВВЕДЕНИЕ

Когда компания небольшая и имеет только один офис, есть возможность проложить физически единую сеть. Однако в случае наличия нескольких офисов даже в пределах одного города, не говоря уже о разных городах и странах, корпоративная информационная сеть (КИС) становится бизнес-инструментом решения для организации и оптимизации трёх бизнес-процессов:

1. Совместная работа с информационными ресурсами. Филиалы, их отделы и сотрудники имеют доступ к единой базе данных, предоставляется возможность единого документооборота и доступа к нему сотрудников.

2. Передача информации. Так как в КИС передаются конфиденциальные данные компании, обеспечение защиты информации важный аспект проектирования КИС. Защита корпоративных сетей обеспечивается как аппаратно, так и программно: межсетевые экраны, виртуальные частные сети, антивирусное программное обеспечение, авторизация и аутентификация, системы обнаружения вторжений (IDS) и предотвращения вторжений (IPS).

3. Деловое общение. В рамках КИС сотрудники, используя специальные программы, могут общаться в чатах, обмениваться файлами и папками, проводить видеоконференции.

Основными свойствами КИС являются:

1. Топология – тип и структура связей между узлами сети.
2. Пропускная способность – возможности сети передавать данные с заданной скоростью.

3. Масштабируемость – создание нового отдела или филиала, например – подключение новых компьютеров должно происходить без значительных изменений в инфраструктуре [1].

Учёт современных требований к безопасности, скорости сходимости и масштабируемости стал важнейшим моментом при проектировании и тестировании корпоративной информационной сети.

ДИНАМИЧЕСКАЯ МАРШРУТИЗАЦИЯ

Топология сети – это, по сути, конфигурация графа, в котором вершинами являются узлы сети (маршрутизатор/компьютер), а рёбрами – связи между вершинами. Зачастую рассматривается не физическая топология компьютеров или других устройств, а логическая, при которой определены пути данных между устройствами в сети. Пути могут задаваться административно – статические маршруты, или – определяться при помощи протоколов маршрутизации – динамическая маршрутизация [2].

Статическая маршрутизация имеет целый ряд недостатков, среди которых основным является плохая масштабируемость: добавление $(N+1)$ -ой сети потребует добавления $2 \cdot (N+1)$ новых записей о маршрутах.

При динамической маршрутизации таблица маршрутизации заполняется и редактируется программно (в UNIX-системах подобные программы называются демонами маршрутизации).

МОДЕЛИРОВАНИЕ СИСТЕМЫ И ЕЁ ТЕСТИРОВАНИЕ

На виртуальном клиенте Linux была развернута лаборатория с помощью утилиты Containerlab, в нее были загружены виртуальные образы ПО межсетевых экранов, маршрутизаторов, серверов и пользовательских станций. Количество межсетевых экранов, соответствовало количеству филиалов, включая центр обработки данных (ЦОД). Отличие от реальной системы лишь в том, что применяется виртуальное оборудование, моделирование при помощи которого эффективнее и дешевле использования реального оборудования: отсутствие затрат на доставку, наладку и поддержку, не требуется место для хранения на сервере, а также лёгкость резервирования.

На границе каждого филиала был настроен Network Address Translation (межсетевые экраны переведены в режим NAT), – механизм, позволяющий преобразовывать IP-адреса транзитных пакетов для удобства и разграничения.

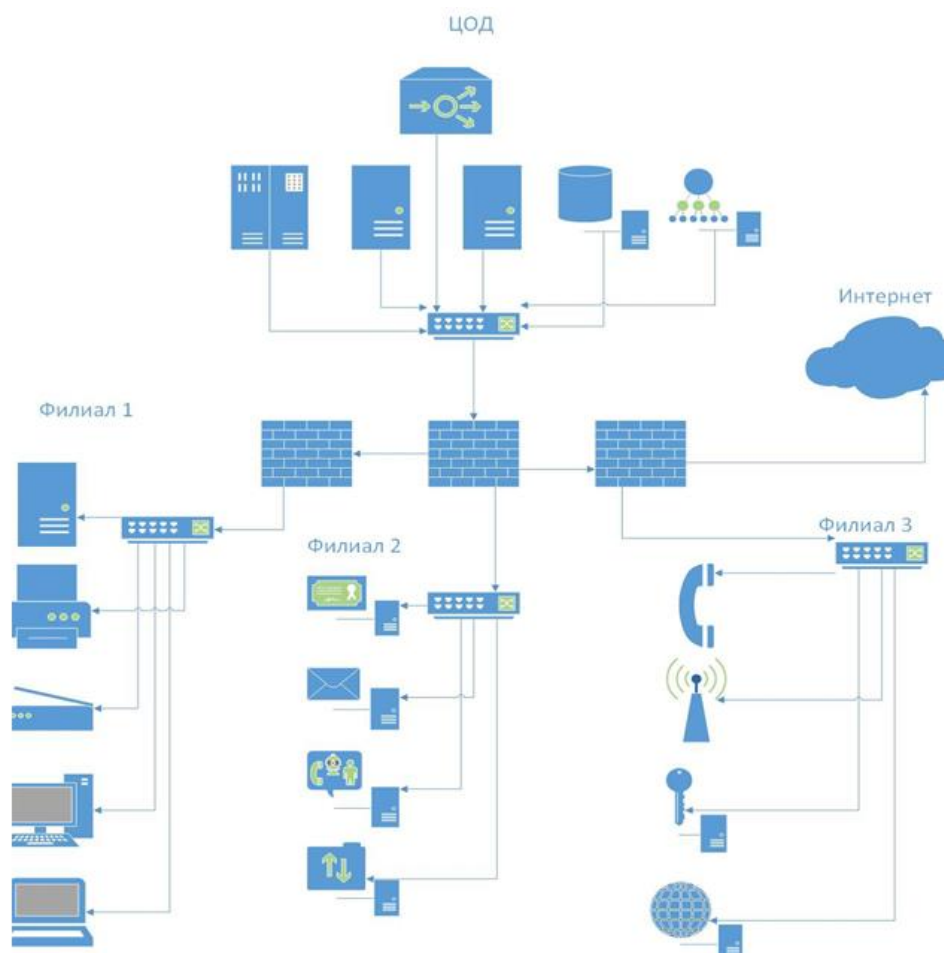


Схема смоделированной корпоративной информационной сети

Маршрутизаторы были настроены так, чтобы внутренняя сеть предприятия функционировала по протоколу внутренней динамической маршрутизации OSPF (Open Shortest Path First – протокол, который составляет карту соседей и прокладывает самый короткий путь из одной точки в другую), а внешняя – по протоколу внешней динамической маршрутизации BGP (Border Gateway Protocol – протокол, обеспечивающий маршрутизацию между автономными системами).

В настоящее время наиболее популярный подход это создание OSPF-зон, между которыми настраиваются статические маршруты [3]. Наша модель включает использование BGP-протокола для организации динамической маршрутизации между OSPF-зонами, а фактически, между несколькими локальными сетями в рамках одной корпоративной информационной сети, что повышает масштабируемость и уменьшает вероятность выхода из строя сети.

В рамках протокола BGP имеется механизм BGP Community, представляющий 32-битовое значение, прикрепляемое к одному или нескольким маршрутам. Такая маркировка помогает при управлении трафиком (возможно указать какой трафик должен быть направлен через конкретные маршруты), группировке маршрутов (создаются группы маршрутов с одинаковыми характеристиками, что упрощает управление).

В каждом филиале трафик перенаправлялся в межсетевой экран, который работал в прозрачном режиме и защищал внутреннюю сеть от вредоносных файлов. На межсетевых экранах были настроены политики для контроля потока трафика, логирование для выявления возможных проблем, внутренние анализаторы для составления статистики и визуализации контроля трафика и обновлена антивирусная база данных, а также настроена маршрутизация для нормального функционирования в системе. Был настроен DNS сервер 8.8.8.8 (Google). Для автоматизации назначения IP-адресов в локальной сети используется *Dynamic Host Client Protocol* с внутренним пулом адресов для автоматической их выдачи соответствующим MAC адресам. На сервере была настроена служба каталогов Active Directory, было создано несколько учётных записей при помощи протокола доступа к каталогам LDAP (*Lightweight Directory Access Protocol*), а также настроен вход по токенам (ключам/тикетам/билетам) через систему аутентификации Kerberos.

ВЫВОДЫ

Таким образом, моделирование корпоративной информационной системы включает проверку редистрибуции двух протоколов динамической маршрутизации (OSPF и BGP) в системе, приближенной к реальной. Используя Eicar-файл, была протестирована безопасность системы, запись отобразилась в логах, файл заблокирован по всем протоколам (FTP, HTTP). По протоколам ICMP (ping) была проверена доступность и сходимость между конечными станциями. Также с помощью трассировки пути было проверено перенаправление трафика на межсетевой экран. С помощью инструментария Containerlab был создан сценарий высоконагруженности системы, сетевого шторма и DOS атаки.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Бройдо В.Л., Ильина О.П. Архитектура ЭВМ и систем: учебник для вузов // 2 изд. Питер, 2021. С. 720.
2. Олифер В., Олифер Н. Компьютерные сети. Принципы, технологии, протоколы // Питер, 2013. С. 944.
3. Дибров М. Сети и телекоммуникации. Маршрутизации в IP-сетях: учебник и практикум для академического бакалавриата // Москва: издательство Юрайт, 2024. С. 423.

КОНЦЕПТУАЛЬНАЯ МОДЕЛЬ И ПРОЕКТИРОВАНИЕ СЛОЖНЫХ ИНТЕГРИРОВАННЫХ СИСТЕМ

В.П. Кочин

*Белорусский государственный университет, Минск, Беларусь
E-mail: kochyn@bsu.by*

Статья посвящена проектированию сложных систем. В статье описывается современное состояние и эволюция процессов проектирования и создания информационных систем. Рассматриваются основные подходы к проектированию сложных систем, основанные на классическом и системном подходах. Описывается подход к проектированию систем, основанный на теории динамического хаоса. На основе результатов анализа существующих подходов к проектированию сложных интегрированных систем обоснована необходимость разработки новых моделей и методов проектирования сложных интегрированных систем. Предложена модель сложных и интегрированных систем, основанная на теории нечетких множеств. Определены основные параметры модели.

Ключевые слова: сложная система; проектирование сложных интегрированных систем; теория нечетких множеств.

Методологии проектирования различных информационных систем развивались с 50-х годов XX века. На начальном этапе, как правило, они относились к локальным информационным системам.

Быстрое развитие в конце 90-х - начале 00-х годов локальных и глобальных сетей, в первую очередь интернета, создало предпосылки для резкого возрастания потребностей в создании различных информационных систем, по сути, процесс их разработки стал носить массовый характер, а сложность возрастать.

Существующие подходы к проектированию сложных информационных систем предназначены для управления различными бизнес-процессами. Как правило, такие системы позволяют управлять отдельными процессами, но при этом имеют интерфейсы для взаимодействия с другими системами. Несмотря на это, данный подход является фрагментарным и не позволяет создавать сложные интегрированные системы управления бизнес-процессами. При небольшом размере системы существующие подходы, которые мы называем, позволяют достаточно эффективно осуществлять цифровую трансформацию бизнес-процессов. В то же время, по мере роста размера системы, фрагментация методов проектирования не позволяет нам проектировать сложные интегрированные системы, позволяющие оптимизировать цифровизацию бизнес-процессов на системной основе [1].

Сложная интегрированная система – это система, состоящая из множества элементов и имеющая многоуровневую структуру как в програм-

мно-техническом плане, так и в организационно-правовом. Многоуровневая структура имеет множество взаимосвязей как между несколькими элементами, так и внутри многоуровневой структуры. В результате сложная интегрированная система приобретает новые функции и возможности, которые не характерны для отдельных элементов системы. Сложная интегрированная система может состоять как из независимых автономных систем, так и из подсистем и других компонентов. Компоненты сложной интегрированной системы могут быть полностью интегрированы в сложную интегрированную систему или частично интегрированы [2]. По определению, сложная интегрированная система состоит из множества элементов, каждый из которых состоит из разных уровней. В зависимости от объекта проектирования количество уровней может варьироваться. В то же время применительно к информационным системам можно выделить основные уровни (рис. 1).

- 1) Технический уровень.
- 2) Программный уровень.
- 3) Организационный уровень.
- 4) Правовой уровень.
- 5) Экономический уровень.

Важной отличительной чертой сложной интегрированной системы является интегрированность, которая определяется наличием множества связей. Причем связи могут быть как между однородными плоскостями различных элементов, так и между разнородными. Количество связей определяется функциями и свойствами сложной интегрированной системы. Например, для реализации определенной функциональности может существовать только одна связь между юридическим уровнем одного элемента и программным уровнем другого элемента. На рис. 1 показана модель сложной интегрированной системы.

Согласно определению, сложная интегрированная система включает в себя множество различных систем и подсистем с различной степенью вхождения на разных уровнях.

$$P = \{P_1; P_2; \dots; P_n\}, \quad (1)$$

где P – сложная интегрированная система;

P_i – элементы (системы, подсистемы), входящие в состав сложной интегрированной системы.

$$P_i = \{O_1^i; O_2^i; O_3^i; O_4^i; O_5^i\}, \quad (2)$$

где O_1^i – технический уровень P_i -го элемента СИС;

O_2^i – программный уровень P_i -го элемента СИС;

O_3^i – организационный уровень P_i -го элемента СИС;

O_4^i – правовой уровень P_i -го элемента СИС;

α_5^i – экономический уровень P_i -го элемента СИС.

Степень вхождения i -го элемента в сложную интегрированную систему P определяется функцией принадлежности

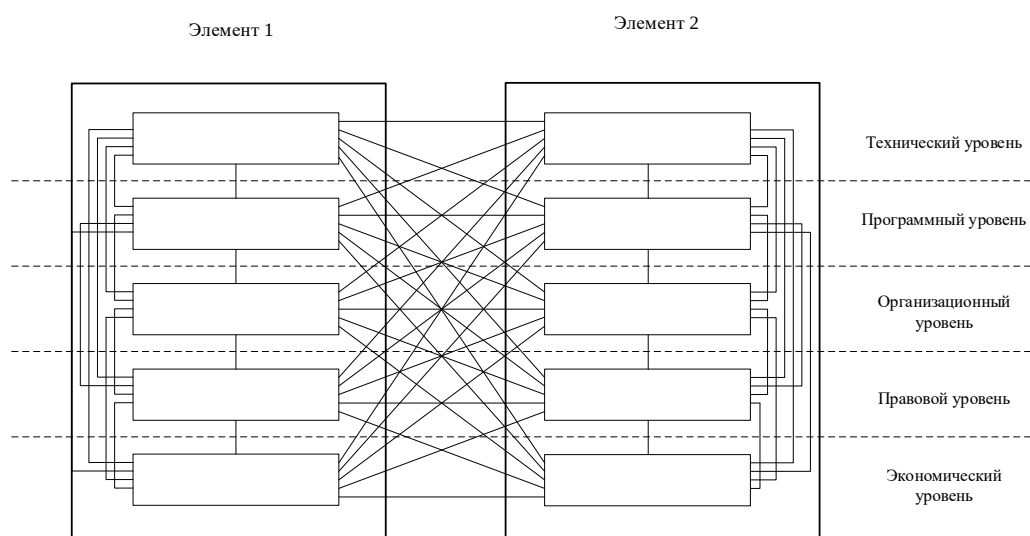
$$P = \{\mu_i(p_i) / p_i\}, \quad (3)$$

где

$$\mu_i : P_i \rightarrow [0,1] \quad (4)$$

характеризует степень принадлежности i -го элемента сложной интегрированной системе P в единичном отрезке от 0 до 1.

μ_i – функция принадлежности i -го элемента множеству P .



Модель сложной интегрированной системы

Таким образом, совокупность функций принадлежности сложной интегрированной системы определяет ее свойства и функциональные возможности. Определение функции принадлежности является одним из важных этапов проектирования сложной интегрированной системы [3].

Одним из этапов проектирования сложной интегрированной системы является задание границ. Для этих целей будут использоваться понятия носителя и высоты сложной интегрированной системы P . Данные характеристики совместно с функцией принадлежности определяют границы сложной интегрированной системы.

Носителем S называется множество всех элементов сложной интегрированной системы P , которые имеют ненулевую функцию принадлежности.

$$S = \{p_i \in P_i : \mu_i(p_i) > 0\} \quad (5)$$

Высота определяет верхнюю границу $\mu_i(p_i)$, т.е. максимальное значение функции принадлежности

$$\sup \mu_i(p_i). \quad (6)$$

Сложная интегрированная система включает в себя различные элементы (системы, подсистемы и др.) с различной степенью вхождения. Таким образом сложная интегрированная система будет определена как:

$$P = \frac{\mu_1(p_1)}{p_1} + \frac{\mu_2(p_2)}{p_2} + \dots + \frac{\mu_i(p_i)}{p_i} = \sum_{i=1}^n \frac{\mu_i(p_i)}{p_i}, \quad (7)$$

где n – количество элементов, входящих в сложную интегрированную систему.

По мере развития информационных технологий менялись подходы к проектированию различных информационных систем. Это было обусловлено ростом сложности информационных систем. Традиционные подходы к проектированию не учитывали рост сложности и количество интеграционных связей. Таким образом, известные подходы к проектированию информационных систем становятся неэффективными для сложных интегрированных систем.

В качестве сложной интегрированной системы может выступать любая социотехническая система, биологическая система и т.д. В рамках данной работы объектом исследования являются информационные системы и платформы различного уровня, а также процессы цифровой трансформации различного уровня [4, 5]. По этой причине исследования проводились применительно к информационным системам и платформам. Для проектирования сложных интегрированных систем предложена формальная модель, основанная на методах нечеткой логики. Данная модель учитывает множество связей как между элементами, так и между уровнями.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Кочин В.П.* Проектирование сложных интегрированных систем как основа успешной реализации стратегии цифровой трансформации // *Веснік сувязі*. 2022. № 5. С. 41–47.
2. *V. Kochyn, A. Kourbatski.* Conceptual design of complex integrated systems // *Открытые семантические технологии проектирования интеллектуальных систем: Сборник научных трудов*, Минск, 2023 г. / Белорус. гос. ун-т информатики и радиоэлектроники; ред-кол.: В. В. Голенков (отв. ред.) [и др.]. Минск, 2023. Вып. 7. С. 33-38.
3. *В.П. Кочин, А.Н. Курбацкий.* Методика проектирования сложных интегрированных систем // *Веснік сувязі*. 2023. № 6. С. 60–63.
4. *V.P. Kochyn, A.V. Zherelo.* Designing a secure fail-safe cloud repository of paperworks of students and employees of educational institutions // *Journal of the Belarusian State University. Mathematics and Informatics*. 2021. № 3. P. 104-108.
5. *В.П. Кочин, Ю.И. Воротницкий.* Проектирование и обеспечение безопасности интегрированных образовательных информационно-коммуникационных систем. Минск: БГУ, 2022. С. 167.

ИССЛЕДОВАНИЕ ПРИМЕНИМОСТИ МЕТОДА СВЯЗАННЫХ ВОЛН ДЛЯ РАСЧЕТА ДИФРАКЦИОННЫХ СПЕКТРОВ ПЕРИОДИЧЕСКИХ РЕШЕТОК СО СЛОЖНЫМ ПРОФИЛЕМ

А.С. Мигель, С.В. Малый

Белорусский государственный университет, Минск, Беларусь

E-mail: anton.migel.99@mail.ru

В работе рассматривается возможность применения метода связанных волн для расчета дифракционного спектра периодической решетки, в виде границы раздела между двумя средами с различными электромагнитными параметрами, профиль которой описывается сложной одномерной функцией. Исследована зависимость точности моделирования от параметров алгоритма, реализующего метод связанных волн.

Ключевые слова: периодическая решетка; дифракционный спектр; метод связанных волн.

Введение. Периодические решетки широко применяются в метрологических системах оптического, инфракрасного и микроволнового диапазонов [1]. Это обусловлено возможностью использования спектра рассеянных на решетке электромагнитных волн в качестве дескриптора ее структуры и материального состава.

Для решения задач дифракции плоских электромагнитных волн на периодических решетках разработаны различные методы, алгоритмы и программы. Среди них можно выделить:

– *универсальные*, основанные на прямом трехмерном или двухмерном моделировании рассеянии электромагнитных волн на периодических решетках (методы конечных разностей во временной области, конечных элементов (МКЭ), конечных интегралов, минимальных автономных блоков [2]);

– *специализированные*, предназначенные для расчета рассеяния электромагнитных волн на решетках (методы связанных волн (МСВ) [3] и Шандезона [4]).

Достоинством первой группы методов является их универсальность. К недостаткам можно отнести высокие требования к вычислительным ресурсам при моделировании большого количества распространяющихся пространственных гармоник. Специализированные методы потенциально обладают большей вычислительной эффективностью, которая зависит от сложности внутренней структуры решеток.

Целью настоящей работы являлось исследование возможности применения метода связанных волн для расчета дифракционного спектра пе-

риодической решетки, в виде границы раздела между двумя средами с различными электромагнитными параметрами, профиль которой описывается сложной одномерной функцией.

Физическая постановка задачи. Рассмотрим взаимодействие плоской линейно поляризованной электромагнитной волны с линейной периодической решеткой в виде границы раздела двух сред, профиль которой описывается сложной функцией одной переменной, как показано на рис. 1. Решетка является бесконечной и периодической вдоль оси x с периодом d . Ось z перпендикулярна плоскости решетки. Для существования распространяющихся гармоник отраженного и прошедшего полей, материалы в соответствующих областях не должны иметь потерь.

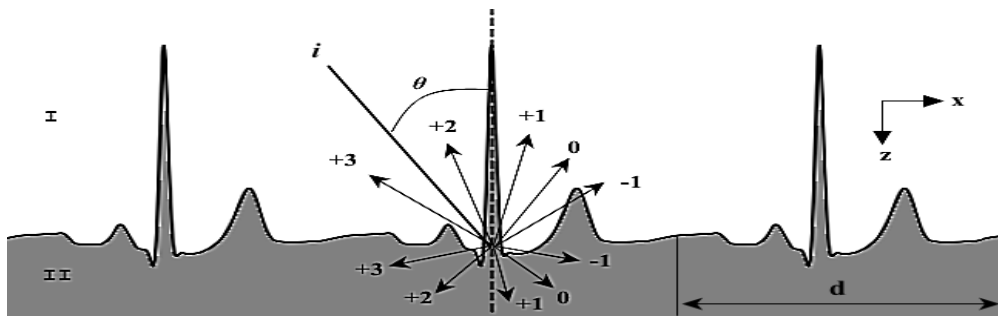


Рис. 1. Геометрия периодической решетки в виде границы раздела двух сред, профиль которой описывается сложной функцией одной переменной

Пусть на решетку из внешней среды под заданным углом θ к оси z падает плоская волна единичной амплитуды ТМ-поляризации (вектор напряженности электрического поля лежит в плоскости падения) с длиной волны λ (направление $\langle i \rangle$ на рис.1). При взаимодействии падающей волны с решеткой происходит дифракция волны на границе раздела сред и формирование конечного числа пространственных гармоник (дифракционных порядков). Для решения задачи используется метод связанных волн.

Результаты моделирования. При использовании метода связанных волн для расчета электромагнитных свойств периодических решеток на точность моделирования оказывают наибольшее влияние количество учитываемых гармоник и количество слоев, описывающих структуру периода. Для оценки точности расчета дифракционного спектра в качестве тестовых значений будут использоваться результаты, полученные с использованием метода конечных элементов и определяемые из условия сходимости решения до 5-го знака.

Пусть профиль периодической решетки описывается функцией вида

$$f(x) = \varphi + \sum_{i=1}^3 A_i \cdot (20 \cdot x - m_i)^2 \cdot \exp(-(20 \cdot x - m_i - \Delta)^2/2), \quad (1)$$

где $\varphi=0,3384$, $\Delta=1$, $A_1=0,0844$, $A_2=0,2815$, $A_3=0,1407$, $m_1=5$, $m_2=11$, $m_3=14$.

На рис. 2 представлен профиль, описываемый выражением (1).

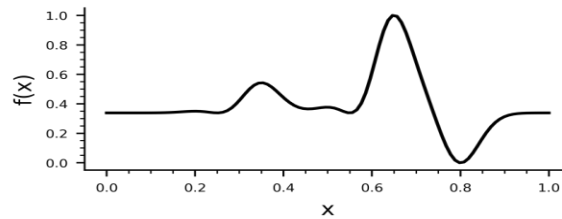


Рис. 2. Профиль периодической решетки

В таблице представлены результаты расчета спектра дифракционных гармоник для различного числа слоев и количества учитываемых гармоник. Нормально падающая на решетку волна имеет ТМ-поляризацию, а граница раздела описывается выражением (1). Длина волны $\lambda=1,0$ мкм. Отношение периода и (высоты) решетки к длине волны равнялось 2,0. Верхнее полупространство – воздух, а нижнее заполнено материалом с диэлектрической проницаемостью равной 2,25.

Спектр дифракционных гармоник, рассчитанный МСВ и МКЭ

Порядок	МСВ						МКЭ
	N=5	N=25	N=50	N=125	N=300	N=500	
Количество учтенных гармоник: 5							
R ₋₁	0,00124	0,00189	0,00201	0,00142	0,00154	0,00152	0,00154
R ₀	0,00948	0,00487	0,00551	0,00541	0,00539	0,00537	0,00533
R ₁	0,0054	0,00368	0,00218	0,00294	0,00284	0,00287	0,00273
T ₋₂	0,04167	0,00585	0,00393	0,00432	0,00429	0,0042	0,00285
T ₋₁	0,02397	0,05774	0,06071	0,05833	0,05935	0,05931	0,06265
T ₀	0,75196	0,76757	0,76253	0,76744	0,76503	0,76519	0,75311
T ₁	0,12733	0,11621	0,11602	0,11238	0,11354	0,11325	0,12283
T ₂	0,03893	0,04219	0,04711	0,04777	0,04801	0,04829	0,04897
Время, с	0,3	0,4	0,6	1,2	1,8	2	30
Количество учтенных гармоник: 25							
R ₋₁	0,00118	0,00184	0,00197	0,00139	0,00152	0,00150	0,00154
R ₀	0,00866	0,00491	0,00542	0,00532	0,00530	0,00526	0,00533
R ₁	0,00573	0,00353	0,00229	0,00296	0,00287	0,00290	0,00273
T ₋₂	0,04075	0,00470	0,00278	0,00327	0,00325	0,00318	0,00285
T ₋₁	0,02771	0,05962	0,06232	0,06014	0,06116	0,06109	0,06265
T ₀	0,73793	0,75937	0,75437	0,75891	0,75653	0,75674	0,75311
T ₁	0,13905	0,12391	0,12286	0,11954	0,12065	0,12035	0,12283
T ₂	0,03900	0,04212	0,04799	0,04846	0,04873	0,04897	0,04897
Время, с	0,5	0,75	1	2	5,5	13	30
Количество учтенных гармоник: 50							
R ₋₁	0,00118	0,00183	0,00196	0,0014	0,00152	0,00151	0,00154
R ₀	0,00866	0,00493	0,00545	0,00532	0,00531	0,00528	0,00533
R ₁	0,00573	0,00353	0,00229	0,00294	0,00285	0,00288	0,00273

Порядок	МСВ						МКЭ
	N=5	N=25	N=50	N=125	N=300	N=500	
T ₋₂	0,04074	0,00465	0,00269	0,00316	0,00315	0,00308	0,00285
T ₋₁	0,02775	0,05969	0,06242	0,06022	0,06124	0,06117	0,06265
T ₀	0,73776	0,75911	0,75381	0,75852	0,75614	0,75637	0,75311
T ₁	0,13918	0,12415	0,12346	0,12014	0,12123	0,12091	0,12283
T ₂	0,03901	0,04210	0,04793	0,04829	0,04857	0,04881	0,04897
Время, с	1	2	3	8	20	44	30
Количество учтенных гармоник: 125							
R ₋₁	0,00117	0,00183	0,00196	0,0014	0,00153	0,00151	0,00154
R ₀	0,00866	0,00493	0,00545	0,00533	0,00532	0,00529	0,00533
R ₁	0,00573	0,00353	0,00228	0,00294	0,00283	0,00286	0,00273
T ₋₂	0,04074	0,00465	0,00268	0,00311	0,00308	0,00302	0,00285
T ₋₁	0,02776	0,05970	0,06243	0,06026	0,06129	0,06122	0,06265
T ₀	0,73775	0,75908	0,75375	0,75832	0,7559	0,75612	0,75311
T ₁	0,13918	0,12420	0,12351	0,12045	0,12159	0,12126	0,12283
T ₂	0,03901	0,04209	0,04792	0,04819	0,04846	0,04872	0,04897
Время, с	3,5	10,4	19	57	160	324	30

Из анализа представленных результатов следует:

- при фиксированном количестве учитываемых гармоник увеличение числа слоев обеспечивает уменьшение погрешности расчета спектра;
- вычислительные затраты на расчет дифракционного спектра в большей степени зависят от количества гармоник, чем от количества слоев;
- количество слоев определяет точность аппроксимации границы раздела (периода решетки).

Заключение. Метод связанных волн может использоваться для расчета спектра распространяющихся пространственных гармоник при рассеянии плоской электромагнитной волны на границе раздела сред, имеющий вид сложной функции одной переменной. Исследование периодических решеток и эффективности применяемых методов, проводилось при финансовой поддержке БРФФИ (проект № Ф23М-043).

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Palmer Ch.* Diffraction Grating Handbook // Richardson Grating Laboratory, New York, eighth edition, 2020.
2. *Малый С. В., Рудницкий А. С.* Методика усредненной матрицы рассеяния в задачах дифракции электромагнитных волн на периодических решетках из металлодиэлектрических лент // Вестник БГУ. 2011. № 2. С. 23-26.
3. *Moharam M. G., Gaylord T. K.* Rigorous coupled-wave analysis of planar-grating diffraction // J. Opt. Soc. Am. 1981. № 71. P. 811 – 818.
4. *Chandezon J., Maystre D., Raoult G.* A new theoretical method for diffraction gratings and its numerical application // J. Opt. (Paris) 1980. № 11. P. 235-241.

ОПРЕДЕЛЕНИЕ CO₂ И ПАРОВ ВОДЫ В АТМОСФЕРЕ

Кугейко М.М., Лынько А.В.

Белорусский государственный университет, Минск, Беларусь

E-mail: kugeiko@bsu.by

Используя базу данных HITRAN проведено моделирование спектра поглощения паров воды и углекислого газа в атмосфере. Получены регрессионные соотношения, связывающие процентное содержание определяемых газовых компонент с главными компонентами, выделенными из спектральных значений их коэффициентов поглощения, для фиксированных значений температур и давления атмосферы, что является основой нового метода контроля парниковых газов атмосферы, позволяющего более эффективно учитывать уширение их спектральных линий даже в условиях сильного перекрытия.

Ключевые слова: коэффициенты поглощения; концентрация газов; перекрытие спектральных линий; регрессионная связь; главные компоненты.

Задача нахождения концентрации газовых компонент из результатов измерений относится к классу некорректных математических. Решение таких задач может быть неустойчивым по отношению к малым вариациям (погрешностям) измеряемых данных, приводит к большим погрешностям в определении компонентного состава смеси. Задача контроля многокомпонентных газовых сред еще больше усложняется когда их спектральные линии сильно перекрываются. В этом случае использование выбранных спектральных каналов измерения, соответствующих линиям поглощения исследуемых компонент, требует учета вклада их перекрытия, что в настоящее время является актуальной задачей.

Ниже рассматривается регрессионный метод решения обратных задач, предполагающий, что искомые параметры среды (в частности, концентрация газовых компонент) могут быть найдены при помощи некоторой статистической функции от измеряемых оптических характеристик. В качестве функциональной связи использовались линейные множественные регрессии: между процентной концентрацией M паров воды и CO₂ и главными компонентами $\varsigma_i, j=1,2$, выделенными из спектральных значений их коэффициентов поглощения $\varepsilon(\lambda_i)$ в интервале 2660-2840 нм с шагом 20 нм:

$$M = a + \sum_i^N b_i \varsigma_i, \quad \varsigma_i = \sum_j^{\Delta\lambda/20} c_{ij} \varepsilon(\lambda_i), \quad (1)$$

где N – количество главных компонент, a и b_i – коэффициенты регрессии, вычисляемые на основе выборки с использованием метода наименьших квадратов [1].

Для решения задачи было проведено моделирование спектра поглощения паров воды и углекислого газа используя базу данных HITRAN [2]. Рассчитывались величины поглощения $\varepsilon(\lambda_i)$ в диапазоне 2640 – 2840 нм с шагом 20 нм для случайных значений концентраций паров H_2O и CO_2 из диапазона от 0,2% до 1% для H_2O и от 0,5% до 2% для CO_2 . Измерительная база задавалась длиной 10 м. При расчете моделируемой выборки $\varepsilon(\lambda_i)$ для каждой температуры и давления учитывалось уширение спектральных линий от их величины. Используя полученные значения $\varepsilon(\lambda_i)$ были выделены главные компоненты ζ_i для каждой из температур и давления, равного 1 атм. Объем каждой из выборок составлял 2500 состояний. Приведённые ниже результаты получены для двух главных компонент ζ_i , которые покрывают 99,9% вариативности данных. Множители при коэффициентах поглощения для главных компонент приведены в табл. 1. Коэффициенты линейной регрессии (1) в табл. 2.

Таблица 1.

Множители при коэффициентах поглощения для главных компонент ζ_i

Длины волн, нм	Значения C_{1i}	Значения C_{2i}
2660	0,3218	-0,2689
2680	0,3263	0,2245
2700	0,2504	0,6195
2720	0,3285	0,1991
2740	0,3155	-0,3217
2760	0,3336	-0,1175
2780	0,3122	0,3449
2800	0,3348	0,0888
2820	0,3152	-0,3232
2840	0,3151	-0,3240

Оценка погрешности восстановления концентраций паров H_2O и CO_2 была проведена на новой выборке $\varepsilon(\lambda_i)$ при наложении на каждое значение из нее погрешности 5% (объем выборки – 500) и составила для паров H_2O и CO_2 соответственно менее двух и трех процентов.

Ранее коэффициенты регрессии и главные компоненты определялись для задаваемого диапазона давления и температуры атмосферы. Установление таких коэффициентов для фиксированных значений температуры и давления является основой нового метода контроля парниковых газов атмосферы даже в условиях сильного перекрытия спектральных линий, позволяющего более эффективно учитывать их уширение, что приводит, при этом, к повышению точности и оперативности метода.

Таблица 2.

Коэффициенты линейной регрессии (1) для H₂O и CO₂.

Т	Коэффициенты линейной регрессии для значений температур для H ₂ O			Коэффициенты линейной регрессии для значений температур для молекул CO ₂		
	a	b1	b2	a	b1	b2
253	-0,01591	0,01548	-0,01927	0,042950	-0,000958	0,025467
255	-0,01283	0,01553	-0,01933	0,038169	-0,000945	0,025525
256	-0,01274	0,01555	-0,01934	0,038805	-0,000941	0,025499
257	-0,01358	0,01559	-0,01936	0,039573	-0,000954	0,025508
258	-0,01103	0,01562	-0,01944	0,034634	-0,000930	0,025545
259	-0,00909	0,01563	-0,01942	0,034975	-0,000943	0,025562
260	-0,00754	0,01570	-0,01953	0,037257	-0,000999	0,025604
261	-0,01278	0,01571	-0,01950	0,031896	-0,000951	0,025655
262	-0,01353	0,01577	-0,01963	0,033959	-0,000966	0,025661
263	-0,01173	0,01580	-0,01961	0,029959	-0,000956	0,025633
264	-0,00969	0,01582	-0,01964	0,031653	-0,000989	0,025703
265	-0,01039	0,01584	-0,01970	0,034543	-0,000956	0,025655
266	-0,01503	0,01590	-0,01975	0,035360	-0,000995	0,025784
267	-0,01173	0,01592	-0,01976	0,031765	-0,000990	0,025735
268	-0,01283	0,01595	-0,01975	0,036097	-0,001000	0,025702
269	-0,00822	0,01598	-0,01993	0,033670	-0,000963	0,025739
270	-0,01015	0,01599	-0,01990	0,034238	-0,000970	0,025791
271	-0,01212	0,01604	-0,01993	0,032681	-0,000982	0,025808
272	-0,00907	0,01605	-0,01996	0,027837	-0,000953	0,025812
273	-0,01186	0,01611	-0,01997	0,036248	-0,000999	0,025724
274	-0,00755	0,01613	-0,02005	0,030166	-0,001010	0,025881
275	-0,01324	0,01613	-0,01999	0,033355	-0,000963	0,025811
276	-0,01425	0,01616	-0,02000	0,038693	-0,000978	0,025806
277	-0,01089	0,01620	-0,02009	0,037693	-0,001002	0,025823
278	-0,01123	0,01623	-0,02012	0,035073	-0,000998	0,025887
279	-0,00892	0,01628	-0,02016	0,030693	-0,001045	0,026008
280	-0,01780	0,01635	-0,02017	0,036686	-0,001040	0,025933

T	Коэффициенты линейной регрессии для значений температур для H ₂ O			Коэффициенты линейной регрессии для значений температур для молекул CO ₂		
	a	b1	b2	a	b1	b2
281	-0,00951	0,01631	-0,02022	0,034097	-0,000987	0,025935
282	-0,00863	0,01636	-0,02033	0,030306	-0,000982	0,025983
283	-0,01002	0,01639	-0,02034	0,033097	-0,001032	0,026098
284	-0,01106	0,01639	-0,02028	0,034419	-0,000995	0,025973
285	-0,00919	0,01646	-0,02036	0,028907	-0,001017	0,026053
286	-0,01394	0,01651	-0,02040	0,031775	-0,001028	0,026046
288	-0,01225	0,01652	-0,02042	0,032759	-0,000993	0,026068
289	-0,01019	0,01657	-0,02051	0,034284	-0,001022	0,026106
290	-0,01387	0,01661	-0,02054	0,036189	-0,001030	0,026151
291	-0,01262	0,01664	-0,02053	0,034649	-0,001051	0,026126
293	-0,00924	0,01668	-0,02066	0,035237	-0,001048	0,026238

Таким образом, для определения количественных значений концентрации каждого компонента газовой смеси предлагаемым методом не требуется сравнений результатов измерения с функцией, описывающей связь регистрируемых сигналов с параметрами исследуемой среды, использования калибровочных сравнительных моделей, устанавливаемых заранее эмпирически (экспериментально), методов решения некорректных обратных задач. При расчете величины поглощения учитываются все перекрытия спектральных линий с учетом их уширения при разных температурах и давлении, что позволяет использовать метод в многокомпонентных средах с перекрывающимися спектральными линиями.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Кугейко М.М., Лысенко С.А. Лазерная спектрофелометрия аэродисперсных сред. Минск: БГУ, 2012. С. 208.
2. Hitranonline [Электронный ресурс]. URL: <https://hitran.org/>.

МОДЕЛИ ОДНОНАПРАВЛЕННОЙ ПЕРЕДАЧИ ДАННЫХ В КОМПЬЮТЕРНЫХ СЕТЯХ

Ю.И. Воротницкий, Р.А. Румас

Белорусский государственный университет, г. Минск, Беларусь

E-mail: vorotn@bsu.by

В докладе рассматриваются модели системы однонаправленной передачи данных для безопасного взаимодействия информационных систем, исключающего несанкционированный доступ, эксплуатацию уязвимостей и иного рода неправомерные действия в отношении источника информации. Для реализации однонаправленной передачи данных в компьютерных сетях предлагается использовать существующий стек протоколов TCP/IP.

Ключевые слова: кибербезопасность; однонаправленная передача данных; оптическая гальваническая развязка.

В современном информационном обществе передача данных играет решающую роль в обеспечении связи и обмена информацией между устройствами и системами. Одной из важных концепций передачи данных является их однонаправленная передача (далее – ОПД), когда информация перемещается от источника к приемнику в строго определенном направлении, исключая возможность атак со стороны, принимающей информацию, на сторону ее передающую. Используемые для этого модели систем однонаправленной передачи данных должны обеспечивать эффективную и надежную доставку информации, минимизируя возможные ошибки и искажения.

В данной работе рассматриваются три основные модели систем однонаправленной передачи данных, их особенности, преимущества и недостатки, области применения. Анализ моделей позволяет лучше понять принципы функционирования систем однонаправленной передачи данных и оптимизировать их производительность в соответствии с требованиями современных коммуникационных задач.

На примере взаимодействия между критически важным объектом информатизации (далее – КВОИ) и корпоративной локальной вычислительной сетью рассмотрим решение проблемы безопасной передачи данных (телеметрии) исключив программно-аппаратное решение – межсетевой экран, не гарантирующий стопроцентную безопасность взаимодействия и не исключающий несанкционированный доступ, в том числе – эксплуатацию уязвимостей информационной системы.

Однонаправленная передача данных предполагает взаимодействие без обратной связи, например, путем использования оптической гальванической развязки [1-3]. Тем самым ОПД гарантируется на физическом уровне, исключая уязвимости и недостатки программного обеспечения

[4]. В случае использования ОПД для повышения безопасного взаимодействия и исключения несанкционированного доступа (сетевых атак) рассмотрим базовую модель, представленную на рис. 1.

Базовая модель ОПД обеспечивает передачу данных (телеметрия, журналы событий и т.п.) от информационной системы с высокой степенью конфиденциальности (в частности, КВОИ) в информационную систему с низкой степенью конфиденциальности (ИС НСК) через аппаратное средство ОПД, функционирующее по принципу оптической гальванической развязки.

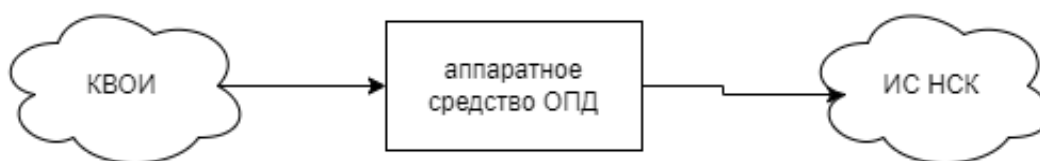


Рис. 1. Базовая модель однонаправленной передачи данных

Преимущества данной модели:

- 1) повышение безопасного взаимодействия при передаче данных, исключая программно-аппаратные СЗИ и используя аппаратное средство ОПД, работающее на физическом уровне;
- 2) доступное аппаратное средство ОПД (например, – комплект оптических медиаконвертеров) для массового применения при передаче данных.

К недостаткам базовой модели можно отнести:

- 1) ограниченный набор функционала из-за работы системы ОПД на физическом уровне и необходимости реализации более высоких уровней модели OSI уже на передающих (принимающих) устройствах (компьютерах, серверах);
- 2) необходимость наличия специального программного обеспечения (далее – СПО), обеспечивающего преобразование работы двунаправленных протоколов FTP, SMB, SMTP и т.п. в однонаправленный поток данных.

В ряде случаев необходимо наличие ОПД, которая подразумевает модель «все включено» (рис. 2). Для решения этой задачи рассмотрим требования к данной модели:

- 1) реализация ОПД на всех уровнях без необходимости использования (установки) СПО на рабочие места пользователей;
- 2) минимизация настройки прокси-серверов, обеспечивающий ОПД, исключая человеческий фактор (ошибки) при их настройке;

3) использование при передаче данных привычных протоколов (FTP, SMB, SMTP и т.д.), которые будут реализованы на прокси-серверах.



Рис. 2. Модель ОПД на основе прокси-серверов

К недостаткам рассматриваемой модели можно отнести:

- 1) наличие значительно большего числа компонентов для реализации по сравнению с базовой моделью и, следовательно, увеличение стоимости системы ОПД;
- 2) необходимость разработки СПО, функционирующего на аппаратных реализациях прокси-серверов;
- 3) отсутствие средств обеспечения отказоустойчивости системы ОПД.

Учитывая недостатки второй модели, можно описать третью модель, включающую следующие требования:

- 1) отказоустойчивая работа по ОПД на стороне получателя (прокси-серверах получателя);
- 2) параллельный прием данных на нескольких прокси-серверах получателя с целью исключения возможных ошибок и проверки контрольных сумм между ними.

Представленная третья модель (рис. 3) хотя и имеет больше компонентов для реализации ОПД, однако снижает вероятность потери сетевых пакетов при передаче одностороннего (например, по протоколу UDP) потока данных за счет наличия двух (или более) прокси-серверов получателя. Прокси-сервера получателя, постоянно ожидая поток UDP данных, готовы принимать и обрабатывать данные. Однако не исключена вероятность сбоя аппаратного или программного обеспечения. Для решения проблемы ошибок приема целесообразно обеспечить «голосование» между

прокси-серверами и проверку (сравнение) полученных данных. Например, при передаче файлов данных могут проверяться передаваемая контрольная сумма и размер файла данных, а при передаче потока UDP трафика (журналов событий, syslog, видеопотока и т.д.) – количество переданных (принятых) пакетов.



Рис. 3. Третья модель однонаправленной передачи данных

В работе рассмотрены три модели ОПД, отличающиеся сложностью реализации, функциональными возможностями и надежностью передачи данных без нарушения их целостности. Выбор одной из моделей определяется конкретными требованиями, предъявляемыми к информационным системам и порядку взаимодействия между ними.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Слепов, Н. Н. Современные технологии цифровых оптоволоконных сетей связи. // М.: Радио и связь, 2003. С. 468.
2. Дмитриев, А. Л. Оптические системы передачи информации // СПб : СПбГУИ-ТМО, 2007.С. 96.
3. Воротницкий Ю.И., Румас Р.А. Оптическая гальваническая развязка при однонаправленной передаче данных // Квантовая электроника: материалы XIV Международной научно-технической конференции, Минск, 21–23 нояб. 2023 г. / Белорус. гос. ун-т ; редкол.: М. М. Кугейко (гл. ред.), А. А. Афоненко, А. В. Баркова. Минск: БГУ, 2023. 1 электрон. опт. диск (CD-ROM). ISBN 978-985-881-530-1. С. 302–305.
4. Ю. И. Воротницкий, Р. А. Румас. Архитектура аппаратно-программного средства однонаправленной передачи данных в компьютерных сетях // Доклады БГУИР. 2023. Т. 21, № 3. С. 96–101. DOI: <http://dx.doi.org/10.35596/1729-7648-2023-21-3-96-101>

ОЦЕНКА КОНЦЕНТРАЦИИ МОЛЕКУЛ В НЕСТАЦИОНАРНОЙ ФЛУОРЕСЦЕНТНОЙ ФЛУКТУАЦИОННОЙ СПЕКТРОСКОПИИ

В.В. Скакун, В.В. Апанасович

*Белорусский государственный университет, Минск, Беларусь;
E-mail: skakun@bsu.by*

Методы флуоресцентной флукутационной спектроскопии позволяют проводить исследования динамики, взаимодействия и структурных изменений молекулярных систем как *in-vitro*, так *in-vivo* путем анализа временных и/или пространственных флукутаций интенсивности флуоресценции. Применение данных методов для анализа нестационарных потоков фотоотсчетов приводят к получению смещенных оценок параметров молекулярных систем. В данной работе, на основе применения теории производящих функционалов, предложен метод оценки концентрации молекул в системе с нестационарным потоком фотоотсчетов. Корректность метода подтверждена на основе анализа смоделированных данных.

Ключевые слова: флуоресцентная флукутационная спектроскопия; поток фотоотсчетов; производящий функционал; производящая функция; концентрация молекул.

ВВЕДЕНИЕ

Флуоресцентная флукутационная спектроскопия (ФФС) это группа методов, позволяющих проводить исследования динамики, взаимодействия и структурных изменений молекулярных систем как *in-vitro*, так *in-vivo*, путем анализа временных и/или пространственных флукутаций интенсивности флуоресценции [1]. Регистрация флукутаций интенсивности флуоресценции производится с однофотонной чувствительностью из предельно малого открытого объема, определяемого лучом лазера, сфокусированного конфокальной оптической системой с высокой числовой апертурой [1]. Методы ФФС разработаны для анализа систем, где зарегистрированный поток фотоотсчетов является стационарным, что накладывает существенные ограничения на спектр их применимости. При проведении измерений в живых клетках редко удается получить измерения со стационарной интенсивностью. Применение же методов ФФС для анализа нестационарных потоков приводят к получению сильно смещенных оценок параметров молекулярных систем, таких как их концентрация и характеристическая яркость, определяемая средним количеством фотонов, излученных одной молекулой в единицу времени. Целью данной работы является разработка метода оценки концентрации молекул в системе с нестационарным потоком фотоотсчетов.

РЕЗУЛЬТАТЫ

Пусть имеется область D , в которой находятся молекулы одного вида, излучающие фотоны. Фотоны регистрируются одноквантовым детектором в течение короткого интервала регистрации и в виде фотоотсчетов (моментов попадания фотонов в детектор) сохраняются в системе. Процесс регистрации повторяется много раз. Поток зарегистрированных фотоотсчетов является потоком случайных событий или случайным точечным процессом, для полного описания которого можно применить аппарат производящих функционалов (ПФЛ) [2]. Обозначим через $\lambda(t, \mathbf{r})$ интенсивность потока фотоотсчетов, зарегистрированных системой от одной молекулы, находящейся в точке с радиус вектором $\mathbf{r}$ в области засветки D в момент времени t . Можно показать, что ПФЛ зарегистрированного потока фотоотсчетов будет иметь вид:

$$L[u; \Omega] = \exp\left\{c \int_D [\exp\{\int_{\Omega} \lambda(t, \mathbf{r}) u(t) dt\} - 1] d\mathbf{r}\right\}, \quad (1)$$

где c концентрация молекул, $\Omega = [T_0, T]$ интервал регистрации, $u(t)$ – формальный параметр (пробная функция) и интегрирование выполняется по всей области D . Достоинством полученного выражения является возможность исследования молекулярных систем с нестационарной интенсивностью флуоресценции, описываемой функцией $\lambda(t, \mathbf{r})$.

Производящая функция (ПФ) числа событий потока имеет вид

$$G(z) = \langle z^n \rangle_n = \sum_{n=0}^{\infty} P(n) z^n, \quad (2)$$

где $P(n)$ – вероятность выпадения на Ω ровно n точек потока и z – некоторая вспомогательная переменная, $0 \leq z \leq 1$. ПФ связана с ПФЛ выражением $G(z) = L[z - 1; \Omega]$. Отсюда следует, что с учетом (1) ПФ потока будет иметь вид

$$G(z) = \exp\left\{c \int_D [\exp\{\int_{\Omega} \lambda(t, \mathbf{r}) (z - 1) dt\} - 1] d\mathbf{r}\right\}. \quad (3)$$

Для стационарных систем

$$\lambda(t, \mathbf{r}) = qB(\mathbf{r}), \quad (4)$$

где q – характеристическая способность молекулы излучать фотоны, равная произведению квантового выхода молекулы, коэффициента экстинкции, эффективности системы регистрации и интенсивности падающего излучения и $B(\mathbf{r})$ функция профиля засветки. Для стационарных потоков можем положить $T_0 = 0$. ПФ потока примет вид

$$G(z) = \exp\left\{c \int_D [\exp\{qTB(\mathbf{r})(z - 1)\} - 1] d\mathbf{r}\right\}, \quad (5)$$

что совпадает с выражением, полученным в работе [3].

Таким образом, получаем полное соответствие с теорией метода FIDA [3], разработанной для анализа стационарных систем.

Для оценки концентрации молекул исследуемого вещества рассчитаем вероятность $P(0)$ отсутствия фотоотсчетов на интервале регистрации Ω . Из теории ПФЛ известно $P(0) = \pi_0(\Omega) = L[u, \Omega]|_{u=-1}$. Тогда

$$\pi_0(\Omega) = \exp\left(c \int_D [\exp\{-\int_\Omega \lambda(t, \mathbf{r}) dt\} - 1] d\mathbf{r}\right). \quad (6)$$

Преобразуем (6) к виду

$$\frac{\ln \pi_0(\Omega)}{c} = \int_D [\exp\{-\int_\Omega \lambda(t, \mathbf{r}) dt\} - 1] d\mathbf{r}. \quad (7)$$

Предположим, что проводим калибровочный эксперимент с известной концентрацией c_k , а затем новый эксперимент с неизвестной c , тогда

$$c = c_k \frac{\ln \pi_0(\Omega)}{\ln \pi_k(\Omega)} = c_k \frac{\ln P(0)}{\ln P_k(0)}. \quad (8)$$

Соответственно, оценку концентрации можно получить на основе выражения

$$\tilde{c} = \tilde{c}_k \frac{\ln \tilde{P}(0)}{\ln \tilde{P}_k(0)}, \quad (9)$$

где $\tilde{P}(0)$ и $\tilde{P}_k$ – оценки вероятностей отсутствия фотоотсчетов за время регистрации и $\tilde{c}_k$ оценка концентрации вещества, полученная в калибровочном эксперименте. Выражение (9) без потери общности можно записать и для среднего количества молекул в области наблюдения, воспользовавшись соотношением $\bar{N} = cV_D$, где V_D характеристический объем области засветки. Полученное выражение позволяет определить концентрацию однокомпонентного вещества на основе калибровочного эксперимента с известной концентрацией и оценок вероятностей отсутствия фотоотсчетов за время регистрации Ω . Как следует из выражения (9), оценку концентрации можно получить для потоков с произвольной (нестационарной) интенсивностью $\lambda(t, \mathbf{r})$ на интервале регистрации Ω без учета функции профиля засветки $B(\mathbf{r})$.

Выполним проверку полученного выражения для оценки концентрации вещества на основе калибровочного эксперимента. Будем использовать имитационную модель потока фототсчетов, разработанную в работе [4], заменив интенсивность потока $I(t)$ на нестационарную

$$I(t) = I_0 e^{-t/\tau_d}, t \in [0; T], \quad (10)$$

где I_0 интенсивность в начальный момент времени T_0 , τ_d константа затухания интенсивности. Параметры модели возьмем следующие: число повторений регистрации 5×10^4 , $q = 10^5$, $T = 2 \times 10^{-5}$ с, $\tau_d = 5 \times 10^{-5}$ с, среднее число молекул в объеме наблюдения для калибровочного эксперимента $\bar{N} = 1$. Получим оценку $\tilde{P}(0)$ для калибровочного эксперимента, затем повторим

моделирование, варьируя $\bar{N}$. Результаты эксперимента приведены в таблице.

Оценка $\tilde{N}$ на основе калибровочного эксперимента

$\bar{N}$ истина	$\tilde{P}(0)$	$\tilde{N}$
1	0,5014	калибровочный
2	0,2512	2,001
5	0,0313	5,017
10	0,0010	9,981
0,1	0,9328	0,101
0,01	0,9932	0,010

Наблюдаем точное восстановление среднего числа молекул в объеме наблюдения в широком диапазоне изменения концентрации молекул, что подтверждает работоспособность разработанного метода.

ВЫВОДЫ

В работе выведено выражение для ПФЛ потока фотоотсчетов, зарегистрированных от некоторой молекулярной системы. Полученный ПФЛ потока открывает возможность исследования молекулярных систем с нестационарной интенсивностью флуоресценции. Определены выражения для расчета вероятности отсутствия фотоотсчетов на интервале регистрации, на основе которого предложен способ оценки концентрации вещества с использованием калибровочного эксперимента. Корректную оценку концентрации можно получить для сильно нестационарных потоков, что характерно, например, при проведении измерений в живых клетках. Справедливость полученного выражения проверена на смоделированных данных.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Confocal-based fluorescence fluctuation spectroscopy with a SPAD array detector / Slen-
ders, E., Castello, M., Buttafava, M. [et al.] // *Light Sci Appl.*, 2021, P. 10, 31. DOI:
<https://doi.org/10.1038/s41377-021-00475-z>.
2. Апанасович В.В. Коляда А.А., Чернявский А.Ф. Статистический анализ случайных
потоков в физическом эксперименте // Минск: Университетское, 1988. С. 256.
3. Kask P. Palo K., Ullmann D. and Gall K. Fluorescence-intensity distribution analysis and
its application in biomolecular detection technology // *Proc. Natl. Acad. Sci. USA*, 1999.
Vol. 96. №. 24. P. 13756.
4. Shingaryov I., Skakun V.V., Apanasovich V.V. Photon counts simulation in fluorescence fluc-
tuation spectroscopy // *Proc. of PRIP conf. Pub. center of BSU, Minsk*, 2009. P. 178-182.

ОНЛАЙН-КАЛЬКУЛЯТОР ДЛЯ РАСЧЕТА ВЫХОДНЫХ ХАРАКТЕРИСТИК ТЕРМОПАР

Д. В. Тиманюк, А.В.Леонтьев

*Белорусский государственный университет, Минск, Беларусь
e-mail: rct.timanyuk@bsu.by*

В работе предоставлен онлайн-калькулятор ThermoType для расчета выходных характеристик всех основных типов термопар. Для разработки выбран язык TypeScript, на основе фреймворка Angular с использованием HCSS препроцесса.

Ключевые слова: температура; термопара; сплав; онлайн калькулятор.

ВВЕДЕНИЕ

Температура – один из основных параметров, который необходимо измерять и контролировать с высокой точностью. Для ее измерения широко используются контактные датчики: термометры сопротивления, термопары и термисторы. В обучающем процессе для изучения методов измерения температуры часто используются онлайн-калькуляторы. Такие приложения значительно упрощают теоретический расчет и исследования в выбранной области, совершенствуют обучающий процесс. Термопара – это наиболее распространенный в промышленности температурный контактный датчик, работающий на основе эффекта Зеебека [1]. Основными преимуществами термопар являются широкий диапазон рабочих температур, возможность непосредственного контакта с поверхностью измеряемого объекта и простота изготовления и использования. Наиболее точными и широко используемыми являются термопары с термоэлектродами из благородных металлов, таких как платинородий-платина (тип S - Pt-10%Rh / Pt) и (тип R - Pt-13%Rh / Pt), а также платинородий-платинородий (тип B - Pt-30%Rh / Pt-6%Rh).

ОСНОВНЫЕ РЕЗУЛЬТАТЫ

Исходя из детального рассмотрения достаточно большого количества онлайн-калькуляторов [2-5] можно сделать вывод, что у каждого сервиса и программного обеспечения присутствуют свои достоинства и недостатки, которые препятствуют их внедрению в учебный процесс. Главным же недостатком является то, что каждая рассмотренная программа имеет потенциальные проблемы с запуском, возможные отсутствия доступа к зарубежному веб-сайту и необходимого дополнительного пакета программ или использование на платной основе.

Таким образом, целью настоящей работы было создание собственного онлайн-калькулятора, который бы позволил в будущем улучшить проведения учебного процесса в соответствующих дисциплинах, а также для личного использования в бытовых целях.

Для разработки выбран язык TypeScript, на основе фреймворка Angular [6] с использованием HCSS препроцессора. Выбор обоснован конечной платформонезависимостью программы, что подразумевает выполнение процессов на различных операционных системах и архитектурах без необходимости перекомпиляции, и современностью инструментов, которые позволяют поддерживать и развивать программу. Блок-схема разработанной программы ThermoType приведена на рис 1.

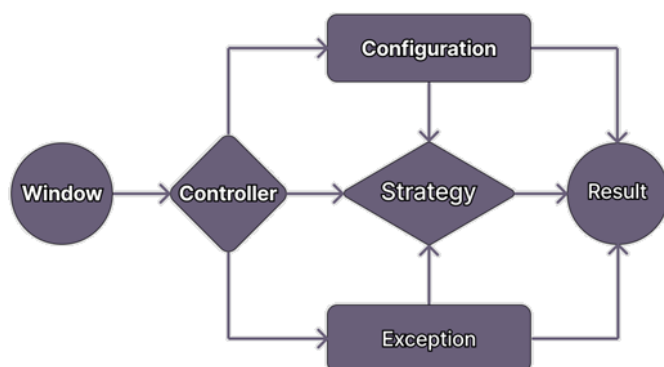


Рис. 1. Блок-схема внутреннего функционала программы ThermoType

Рабочее окно программы приведено на рис.2.

Рис. 2. Рабочее окно разработанной программы ThermoType

При переключении в основное окно браузера, начинаем работу с калькулятором ThermoType. После ввода данных отправляется запрос на сервер, то есть в пакет controller, если данные введены правильно, то система направляет в стратегию, если нет то в exception – работа прекращается.

Программа ThermoType позволяет работать с основными типами термопар, приведенными в ГОСТ Р 8.585 -2001: R, S, B, J, T, E, K, N, A-1, A-2, A-3, L, М. Для расчета выходных характеристик термопар используются полиномы из указанного ГОСТа. Предлагается выбор единиц измерения температуры. Если требуется перевести температуру в ТЭДС, вводим температуру, если из ТЭДС в температуру, то наоборот. Важной особенностью разработанного калькулятора является опция задания температуры холодного спая, которая может устанавливаться в пределах от 0 до 60°C. Далее результат обрабатывается в пакете ORIGIN.

ЗАКЛЮЧЕНИЕ.

В ходе проведенного исследования был разработан алгоритм расчёта характеристик термопар на базе языка TypeScript с использованием фреймворка Angular, с учетом температуры холодного спая. Проведенные тестовые расчет температуры и ТЭДС по программе ThermoType полностью согласуются с данными ГОСТа Р 8.585-2001 и хорошо соотносятся с результатами расчётов одного из популярных онлайн калькуляторов [5]. Разработанный онлайн калькулятор допускает использование операционных систем Windows и Android, что позволяет проводить удаленные лабораторные работы.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Термоэлектрические преобразователи температуры. Теория, практика, развитие / Под общ. ред. А. В. Каржавина. Обнинск: ПК «Тесей», 2004 г. С. 84.
2. Онлайн-калькулятор для расчёта выходных параметров термопар [Электронный ресурс]. URL: <https://rcl-radio.ru/?p=57835> (дата обращения: 27.03.2024).
3. Онлайн-калькулятор для расчёта выходных параметров термопар [Электронный ресурс]. URL: <https://us.flukecal.com/Thermocouple-Temperature-Calculator>. (дата обращения: 05.04.2024).
4. Онлайн-калькулятор для расчёта выходных параметров термопар [Электронный ресурс]. URL: <https://gock.net/thermocouple-converter> (дата обращения: 05.04.2024).
5. Онлайн-калькулятор для расчёта выходных параметров термопар [Электронный ресурс]. URL: <https://australianoxytrolsystems.com/calculators/thermocouple> (дата обращения: 05.04.2024).
6. Документация к программному пакету Angular [Электронный ресурс]. URL: <https://angular-doc.ru/guide/architecture> (дата обращения: 27.03.2024).

OPTIMAL CONTROL SYSTEM STRUCTURE FORMATION OF HUMAN-MACHINE SYSTEMS USING GENETICAL ALGORITHMS

Huang Ruiqi, A.A. Bezrodny

Belarusian State University, Minsk, Belarus

E- mail: Ricky867626073@gmail.com

In the advancing technological era, optimizing human-machine systems is crucial for enhancing performance and user experience. The complexity of such systems and the unpredictability of human behavior present significant challenges. Genetic algorithms (GAs), inspired by biological evolution, offer a proficient method for addressing these issues by simulating natural selection processes. This enables the effective exploration of solution spaces, identifying optimal control structures for improved system performance and adaptability across various scenarios. GAs' flexibility further allows for effective solutions in diverse applications like aerospace, autonomous vehicles, and manufacturing, improving human-computer interaction.

Key words: Human-machine system optimal control; genetic algorithm; solution space; adaptability; causal approach.

ФОРМИРОВАНИЕ ОПТИМАЛЬНОЙ СТРУКТУРЫ СИСТЕМЫ УПРАВЛЕНИЯ ЧЕЛОВЕКО-МАШИННЫЕ СИСТЕМЫ С ИСПОЛЬЗОВАНИЕМ ГЕНЕТИЧЕСКИХ АЛГОРИТМОВ

Хуан Жуйци, А.А. Безродный

Белорусский государственный университет, Минск, Беларусь

e-mail: Ricky867626073@gmail.com, bezrodnyaa@yandex.ru

В развивающуюся технологическую эпоху оптимизация человеко-машинных систем имеет решающее значение для повышения производительности и удобства пользователей. Сложность таких систем и непредсказуемость поведения человека создают серьезные проблемы. Генетические алгоритмы (ГА), вдохновленные биологической эволюцией, предлагают эффективный метод решения этих проблем путем моделирования процессов естественного отбора. Это позволяет эффективно исследовать пространство решений, определяя оптимальные структуры управления для повышения производительности системы и ее адаптируемости в различных сценариях. Гибкость ГА также позволяет создавать эффективные решения в различных приложениях, таких как аэрокосмическая промышленность, автономные транспортные средства и производство, улучшая взаимодействие человека и компьютера.

Ключевые слова: оптимальное управление человеко-машинной системой; генетический алгоритм; пространство решений; адаптивность; причинный подход.

INTRODUCTION

This study employs Genetic Algorithms (GA), grounded in evolutionary principles, to tackle the complex optimization challenges of controlling Cyber-

Physical Systems (CPS) [1], which integrate computational algorithms and physical processes. These systems play a pivotal role in areas such as smart cities, healthcare, and manufacturing. By adopting a causal approach [2], we systematically explore the solution space for CPS control, aiming to identify optimal configurations that balance performance with adaptability. This approach emphasizes understanding the interactions between computational and physical elements, with the goal of enhancing the efficiency and reliability of CPS.

The goal of this study is to optimize the control structure of human-machine systems using genetic algorithms, in order to solve the problems of redundancy and inflexibility in current designs, thereby improving the efficiency of human-machine interaction (in the case of energy supply). To achieve the goal, the following tasks shall be solved:

1. Analyze existing research on the formation, improvement, and evaluation of control system structures to identify gaps in scope, methodology, and unexplored areas.
2. Select the most suitable software tool to construct algorithm models and workflows. Initially employing a causal approach, with the possibility of later adjustments to a more effective method.
3. Test, revise, and implement the results within petroleum supply companies.

GA MODELLING

To make the modeling more intuitive, we will use a petroleum supply company as an example for simple modeling. The design structure diagram to be incited in the modelling process is shown in fig. 1.

Allocation ratios of raw materials in refining engineering are $p_1, p_2, \dots, p_n$, where p_i represents the ratio of the i type of crude oil in the refining process. Terminal logistics management are $c_1, c_2, \dots, c_m$, representing the number of terminals. Transport route selection: Use binary variables t_{ij} to indicate whether to choose the transport route from location i to location j , where $t_{ij} = 1$ indicates selection, and $t_{ij}=0$ indicates no selection. Supply frequency and quantity of gas stations are $s_1, s_2, \dots, s_k$, representing the supply frequency of gas stations, mapped to the integer range of $[1, 7]$; $a_1, a_2, \dots, a_L$, representing the supply quantity of gas stations, in units of kiloliters.

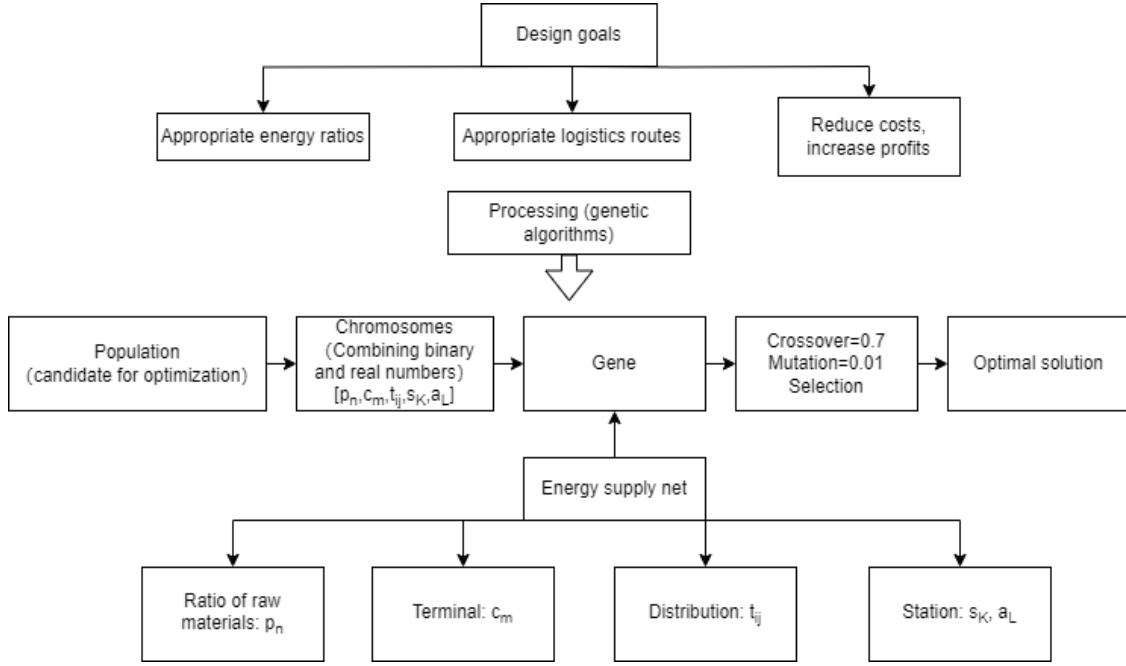


Fig. 1. Design structure diagram

Objective function (also fitness function) is defined as

$$\max_{P, C} Z = \alpha \times P - \beta \times C, \quad (1)$$

where P – total production, where C is the total cost, where α and β are the weight coefficients.

Constraints, $\sum_{i=1}^n p_i = 1$, make sure all ratios notmalized to 1. The value range of raw material distribution ratio is $0 \leq p_i \leq 1$. The supply quantity is a non-negative number. There must be at least one terminal and one route selected.

To make a model the Roulette Selection method has been used. Roulette selection is a stochastic selection method, see in fig. 2, where the probability for selection of an individual is proportional to its fitness. The method is inspired by real-world roulettes. That means, however, that all slots have the same probability of being selected. Instead, we can implement a weighted version of the roulette. With it, the larger the fitness of an individual is, the more likely is its selection. In a population with n individuals, for each chromosome x with a corresponding fitness value f_x , we compute the corresponding probability p_x of selection (eq. (2)).

$$p_x = \frac{f_x}{\sum_{i=1}^n f_i} \quad (2)$$

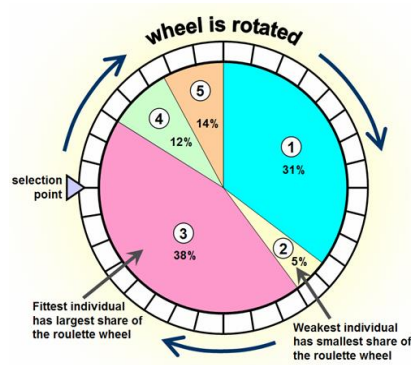


Fig. 2. Roulette Selection

In the next step we have to define the coding strategy. For optimizing a petroleum supply chain that includes refining, logistics, transportation, and supply to gas stations, the coding uses real numbers for raw material distribution ratios (e.g. 0.3, 0.2, 0.5), binary coding for logistics management and route selection decisions ([1, 0, 1] for choosing specific logistics strategies and [1, 0, 0, 1] for selecting routes), and real numbers again for the frequency and quantity of gas station supplies, like [5, 40] to indicate 40 kiloliters supplied five days a week.

SIMULATION RESUTLS

Suppose we have a medium-sized petroleum company that faces a supply problem in the following situation, see tables 1 and 2.

Table 1

Crude oil prices and refining efficiency

	Price	Refining efficiency
P1	20 USD/barrel	80%
P2	15 USD/barrel	70%
P3	30 USD/barrel	85%

Refining cost of 5 USD/barrel of crude oil and product selling price 100 USD/barrel.

Table 2

Terminal management costs and transportation costs

	C1: 1,000 USD/day	C2: 900 USD/day
G1	200 USD/barrel	200 USD/barrel
G2	150 USD/barrel	150 USD/barrel

Set the parameter values as $N = 100$, $\eta_{\text{mut}} = 0.01$, $\eta_{\text{cross}} = 0.7$, $Gene = 100$. The results of the calculations are [0.598, 0.131, 0.271, 0, 1, 0, 1, 0, 0, 7, 3, 21.43, 83.3]. The maximum profit is \$1284.43.

CONCLUSION

This research highlights genetic algorithms effectiveness in enhancing human-machine system controls, particularly for a petroleum supply company. GAs helped overcome design challenges like redundancy and inflexibility, significantly improving system efficiency, adaptability, and user experience. Future directions include exploring GA implementations in real scenarios, examining human factors like user acceptance, and employing more complex models (increasing the number of terminals, adding KPIs, etc.) to deepen our understanding.

BIBLIOGRAPHY

1. *Alexey Bezrodny, Anatoliy Korolenok, Aleksandr Rezhnikov*. Complex human-machine petroleum system improve by causation approach // Open Semantic Technologies for Intelligent Systems. Research Papers Collection. Issue 5. Minsk, Belarus, 2021. P. 251-256.
2. *Goldberg, D. E.* Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Longman Publishing Co., Inc. 1989.
3. *Fisher, R. A.* The genetical theory of natural selection. Oxford: Clarendon Press. 1930.
4. *Kristinsson, K., & Dumont, G. A.* System identification and control using genetic algorithms. Journal of Process Control, 2(4), 1992. P. 189-204.
5. *Oliveira, Paulo, João Sequeira, and Joao Sentieiro*. "Selection of controller parameters using genetic algorithms." Engineering Systems with Intelligence: Concepts, Tools and Applications, 1991. P. 431-438.
6. *Qian Zhiqin, Teng Hongfei, Sun Zhiguo*. Genetic Algorithms in Human-Machine Interaction and Their Application in Constrained Layout Optimization[J]. Journal of Computer Science, 2001. (05): P. 553-559.

АППАРАТНО-ПРОГРАММНАЯ ПЛАТФОРМА ИССЛЕДОВАНИЯ АЛГОРИТМОВ МАРШРУТИЗАЦИИ В БЕСПРОВОДНОЙ СЕНСОРНОЙ СЕТИ

А.К. Шульган, Ю.И. Воротницкий

*БГУ, Минск, Республика Беларусь
E-Mail: lightworksby@gmail.com*

В статье предлагается аппаратно-программная платформа для моделирования процессов сбора информации, маршрутизации и мониторинга процессов в беспроводных сенсорных сетях.

Ключевые слова: сенсорные сети; многокритериальная маршрутизация; технология Zigbee

На протяжении последних лет области применения беспроводных сенсорных сетей расширяются, включая в себя автоматизацию зданий (мониторинг присутствия и поведения людей, управления освещением, климатическими установками и энергоснабжением, мониторинг состояния несущих конструкций зданий и т.п.), промышленную автоматизацию (системы телеметрии, диагностика оборудования, мониторинг производственных процессов и др.), здравоохранение, сельское хозяйство, системы безопасности. Пространственно распределенные сети состоят из датчиков (сенсоров) и исполнительных устройств, объединенных между собой посредством радиоканала. Сенсорные сети имеют свойство самоорганизации: устройства в сети способны к автономному нахождению маршрута друг к другу, а также к центральному узлу. В таких сетях в последнее время все чаще используются новые для сенсорных сетей типы трафика, в том числе – видео- и аудио- трафик. Формулируются новые требования к телекоммуникационным параметрам, в частности, к параметрам качества обслуживания. Оптимизация беспроводных сенсорных сетей может осуществляться как путем поиска оптимальных маршрутов доставки сетевых пакетов, так и путем оптимизации размещения узлов и вариации их параметров. Алгоритмы маршрутизации, применимые для мультисервисных сенсорных сетей, рассмотрены в [1,2]. Вопросы качества обслуживания в подобных сетях исследовались в [3].

В настоящей работе предложен аппаратно-программный комплекс для исследования алгоритмов маршрутизации и параметров качества обслуживания в беспроводных сенсорных сетях. В нем используется широко распространенная технология беспроводной связи в диапазоне 2,4 ГГц Zigbee, предназначенная для организаций распределённых, самоорганизующихся и малопотребляющих сетей. Физический слой и управление доступом к среде для

таких сетей описаны в стандарте IEEE 802.15.4, который определяет скорость передачи данных до 250 Кбит/с и топологию Mesh.

Моделирование предполагает возможность реализации сетей с различной топологией (рис. 1). При этом узлы могут представлять конечные устройства, специализированные узлы-маршрутизаторы и узлы-координаторы. В ряде случаев узлы сенсорной сети могут выполнять несколько перечисленных функций.

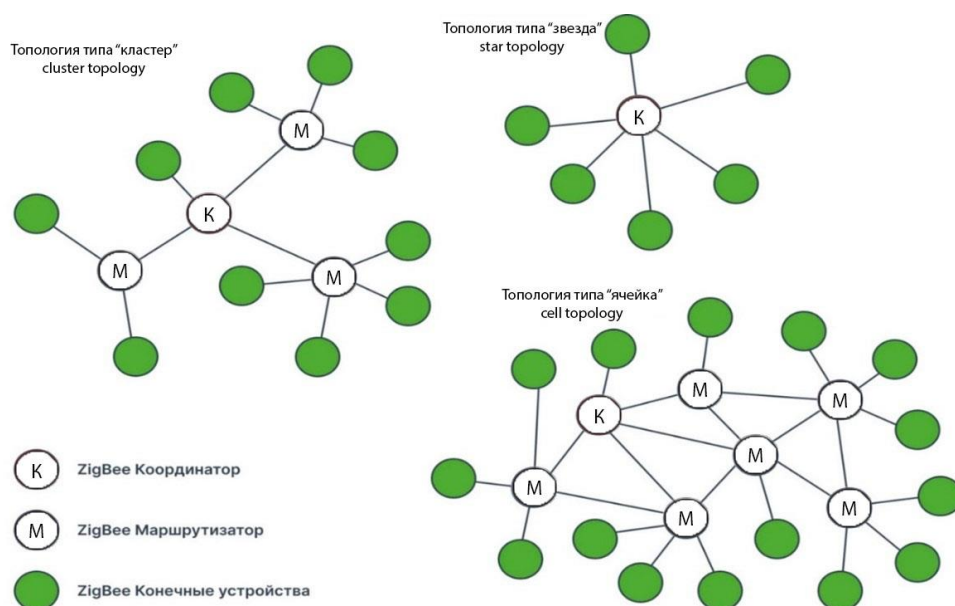


Рис. 1. Различные топологии сетей Zigbee

Для исследования процесса построения маршрута, исходя из параметров узлов, была разработана модель беспроводной сенсорной сети. Данная сеть состоит из нескольких многофункциональных узлов, построенных на базе доступных и широко распространенных микроконтроллеров ATmega328. Протоколы Zigbee были реализованы с помощью радиомодемов XBee S2C. Для настройки радиомодемов использовались адаптеры XBee Explorer, с интерфейсом USB. Схема реализованного в настоящее время макета показана на рис. 2.

В текущей конфигурации сеть состоит из узла-координатора, трёх узлов, выполняющих роль маршрутизаторов, а также конечных устройств – исполнительных устройств и датчиков, входящих в состав систему «умного дома». Узел-координатор подключён к персональному компьютеру, являющемуся устройством сбора информации и управления. Для разработки приложений, устанавливаемых на узлах, предлагается использовать TinyOS и Contiki-ng. Приложения обеспечивают функционирование сети,

выполняя мониторинг ее параметров и построение оптимальных маршрутов и реализуя основные подходы к оптимизации параметров беспроводных сенсорных сетей, предложенные в [4].

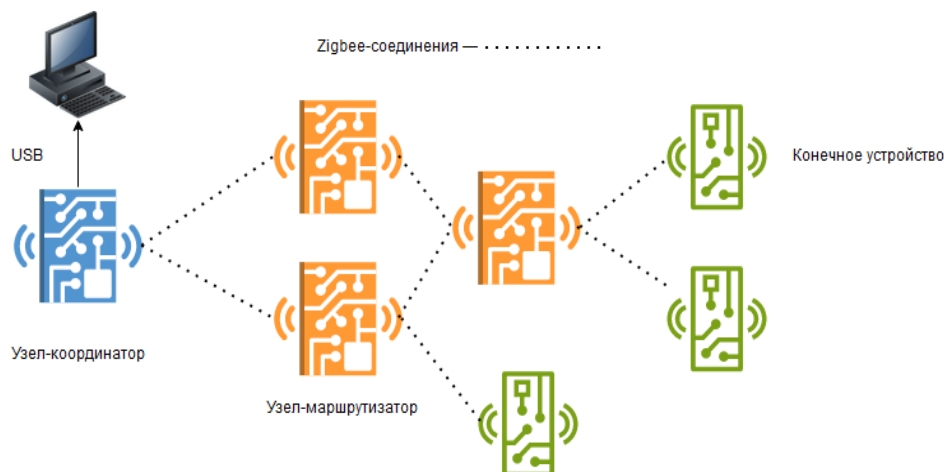


Рис. 2. Принципиальная схема макета беспроводной сенсорной сети

В результате можно сделать вывод о возможности моделирования процессов оптимальной маршрутизации в беспроводных сенсорных сетях на аппаратной платформе ATmega328 с радиомодемами XBee S2C.

В настоящее время предметом исследования с использованием разработанной платформы являются алгоритмы многокритериальной маршрутизации в сенсорной сети, обеспечивающей, в том числе, сбор и передачу потоковой видео и аудио информации по протоколу Zigbee.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Листопад Н. И., Воротницкий Ю. И., Бортновский В. В., Хайдер А. А. Многокритериальная маршрутизация информационных потоков // Проблемы физики, математики и техники. 2017. №2(31). С.84-90.
2. Листопад Н. И., Воротницкий Ю. И., Хайдер А. А. Маршрутизация в мультисервисных сетях телекоммуникаций на основе модифицированного алгоритма Дейкстры // Вестник БГУ. Сер.1, Физика. Математика. Информатика. 2015. №1. С.70-76.
3. Мулярчик К. С., Полочанский А. С. Качество обслуживания в беспроводных сенсорных сетях // Журн. Белорус. гос. ун-та. Математика. Информатика. 2017. № 2. С. 65–70.
4. Ю.И.Воротницкий, А.К. Шульган. Анализ подходов к оптимизации параметров беспроводных сенсорных сетей // Материалы Международной научно-практической конференции «Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния», Минск, 18–19 мая. 2023 г. Минск: Научно-исследовательское учреждение «Институт прикладных физических проблем имени А.Н. Севченко», 2023. С. 146-148.

ГЕНЕРАТИВНОЕ ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ СЛОЖНЫХ БИОФИЗИЧЕСКИХ СИСТЕМ

Н.Н. Яцков, В.В. Апанасович, В.Н. Яцков

Белорусский государственный университет, Минск, Беларусь

E-mail: yatskou@bsu.by

Разработан метод генеративного имитационного моделирования для исследования биофизических систем, информация о поведении которых представлена в виде больших наборов экспериментальных данных. Предложенный метод является инструментом искусственного интеллекта и основан на генерации адекватных сценариев поведения систем, используя имитацию случайных событий, вероятностные распределения которых находятся из имеющихся экспериментальных данных. Применение метода позволяет повысить эффективность анализа сложных биофизических систем.

Ключевые слова: искусственный интеллект; генеративное моделирование; имитационная модель; машинное обучение; биофизическая система.

ВВЕДЕНИЕ

Использование методов имитационного моделирования в ходе анализа больших данных открывает новые возможности в области высокотехнологичных междисциплинарных исследований [1]. Имитационное моделирование позволяет осуществлять анализ систем практически любой сложности и проводить вычислительные эксперименты, труднодостижимые в натуральных исследованиях [1].

В современных научных исследованиях наблюдается тенденция применения генеративного моделирования, целью которого является создание реалистичных и разнообразных образцов данных по заданным распределениям [2]. Генеративная модель описывает, как генерируется набор данных, с точки зрения вероятностной модели. Используя эту модель, можно генерировать новые данные [3].

Генеративное моделирование может использоваться в биоинформатике для проектирования новых белков, молекул, геномов и клеток, а также для изучения основных механизмов и закономерностей биологических систем [4, 5]. Отдельное направление применения генеративного моделирования связано с формированием обучающих данных для методов машинного обучения с целью анализа данных реальных экспериментов [6]. В этом случае формирование смоделированных обучающих данных может иметь преимущества по точности и эффективности при анализе экспериментальных данных, измеренных в различных условиях, обусловленных экспериментальными искажениями. Предполагается, что обуче-

ние на смоделированных данных конкретного эксперимента позволит повысить точность алгоритмов машинного обучения при анализе биофизических систем [7].

В работе предлагается метод генеративного имитационного моделирования для исследования сложных биофизических систем.

ГЕНЕРАТИВНОЕ ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ

Предполагается, что через серию экспериментов, порождающих наборы больших данных, исследуется некоторый биофизический объект E , существенные признаки, или характеристики которого A , должны быть определены в ходе анализа данных. Обозначим через $H_E = \{E_1, E_2, \dots, E_N\}$ набор данных, состоящий из множества наблюдений, примеров или измерений объекта E , образы или новые проявления которых необходимо сгенерировать. Каждое наблюдение состоит из множества свойств или признаков двух видов: первые (обозначим их X) представляют собой независимые измерения, вторые (обозначим их Y) представляют собой измерения в зависимости от выбранных значений признаков первой группы. Измерения над объектами формируют векторы признаков $X_1, X_2, \dots, X_K$ и матрицу входных данных X . Цель генеративного моделирования – создать модель, способную генерировать новые наборы признаков ($\dot{X}, \dot{Y}$), которые выглядят так, будто созданы с использованием тех же правил, что и исходные данные.

Генеративная модель должна быть вероятностной и включать элементы, которые влияют на отдельные выборки, генерируемые моделью. Можно говорить о том, что существует какое-то неизвестное вероятностное распределение $f(X)$, характеризующее исследуемую систему. Основная задача — создать модель, максимально точно имитирующую распределение $f(X)$, а затем произвести выборку из нее, чтобы сгенерировать новые наблюдения, которые выглядят так, будто могли бы быть в исходном обучающем наборе.

Предполагается, что выполняется анализ больших данных (X, Y) , наблюдения H_E экспериментально измерены или сгенерированы в соответствии с некоторым неизвестным распределением $f(X, \Theta) = \{f_1(X_1, \Theta_1), f_2(X_2, \Theta_2), \dots, f_K(X_K, \Theta_K)\}$, где $f_i(X_i, \Theta_i)$ – функция плотности распределения для признака X_i , Θ_i – вектор параметров распределения f_i . В общем случае решаются задачи *классификации* и *регрессии* [1].

Обобщенный алгоритм генеративного имитационного моделирования.

Шаг 1. Выполняется анализ признаков X с целью выявления законов распределений $f_1(X_1, \Theta_1), f_2(X_2, \Theta_2), \dots, f_K(X_K, \Theta_K)$ и последующей оценкой

их параметров Θ . Строятся гистограммы $h_1, h_2, \dots, h_K$ признаков $X_1, X_2, \dots, X_K$. Гистограммы аппроксимируются функциями $f'_1(X_1, \Theta'_1), f'_2(X_2, \Theta'_2), \dots, f'_K(X_K, \Theta'_K)$, приближениями законов распределений $f_1(X_1, \Theta_1), f_2(X_2, \Theta_2), \dots, f_K(X_K, \Theta_K)$. С использованием метода оптимизации оцениваются параметры Θ' .

Шаг 2. Восстановленные законы распределений $f'(X, \Theta')$ и оценки их параметров Θ' используются в имитационной модели M исследуемых физических характеристик объекта. Имитационная модель генерирует новые данные $(\dot{X}, \dot{Y})$. Правильно подобрав имитационную модель обобщенного алгоритма генеративного имитационного моделирования комплексных биофизических систем, а именно – законы распределений $f'_i(X_i, \Theta'_i), i = 1, 2, \dots, K$, можно генерировать новые наблюдения, характеризуемые набором данных $(\dot{X}, \dot{Y})$, которые выглядят так, будто были получены из распределения $f(X, \Theta)$.

Шаг 3. Строятся модели машинного обучения $m = \{m_1, m_2, \dots, m_l\}$ для последующего анализа экспериментальных данных. Смоделированные данные $(\dot{X}, \dot{Y})$ используются для обучения алгоритмов m . Оценивается предсказательная точность моделей машинного обучения.

Шаг 4. Обученные на смоделированных данных алгоритмы машинного обучения m применяются к анализу экспериментальных данных (X, Y) . Определяются параметры A .

АДЕКВАТНОСТЬ ГЕНЕРАТИВНЫХ ИМИТАЦИОННЫХ МОДЕЛЕЙ И АНАЛИЗ ТОЧНОСТИ МОДЕЛИРОВАНИЯ

Будем полагать генеративную имитационную модель *адекватной* – если модель с заданной точностью воспроизводит восстановленные законы распределений $f'(X, \Theta')$ и *точной* – если обученные на смоделированных данных модели машинного обучения m оценивают неизвестные параметры A объекта E не хуже или лучше, чем классические алгоритмы анализа данных.

ПРИМЕНЕНИЕ ГЕНЕРАТИВНОГО ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ К АНАЛИЗУ БИОФИЗИЧЕСКИХ СИСТЕМ

Проверка адекватности разработанных моделей и анализ точности моделирования выполнены на примерах решения задач *классификации* – идентификации сайтов однонуклеотидных полиморфизмов и *регрессии* – предсказания выживаемости онкопациентов на примерах больших наборов данных геномного секвенирования. Полученные результаты позволяют сделать вывод о том, что для реальных экспериментальных данных предпочтительнее использовать модели машинного обучения, обученные

на смоделированных данных. Средняя точность идентификации сайтов однонуклеотидных полиморфизмов на 2-5% выше для моделей машинного обучения, чем для классических статистических методов. При применении генеративного имитационного моделирования для обучения моделей предсказания выживаемости достигнуто почти двукратное снижение ошибки предсказания.

ЗАКЛЮЧЕНИЕ

Разработан метод генеративного имитационного моделирования для исследования сложных биомолекулярных систем по экспериментальным наборам данных, основанный на генерации случайных событий вероятностных моделей распределений, параметры которых оцениваются по имеющимся экспериментальным данным, и методов машинного обучения, обученных на смоделированных данных и применяемых при решении задач *классификации и регрессии*. Использование алгоритмов машинного обучения способствует более точному определению закономерностей, получаемых в результате анализа больших экспериментальных данных. Применение разработанного метода позволяет повысить эффективность анализа при идентификации сайтов однонуклеотидных полиморфизмов и предсказании выживаемости онкопациентов по данным геномного секвенирования.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Яцков Н. Н., Апанасович В. В. Комплексный анализ данных при исследовании сложных биомолекулярных систем // Информатика. 2021. Т. 18, № 1. С. 105–122. DOI: <https://doi.org/10.37661/1816-0301-2021-18-1-105-122>
2. Generative Adversarial Networks and Deep Learning / 1st ed. Eds.: R. Raut, P. D. Pathak, S. R. Sakhare, S. Patil. Boca Raton: Chapman and Hall/CRC, 2023. P. 208.
3. Foster D. Generative Deep Learning : Teaching Machines to Paint, Write, Compose, and Play / 2nd ed. Oreilly & Associates Inc., 2023. P. 456.
4. Ruan X., Murphy R. F. Evaluation of methods for generative modeling of cell and nuclear shape // Bioinformatics. 2019. Vol. 35(14). P. 2475-2485. DOI: 10.1093/bioinformatics/bty983.
5. Kucera T., Togninalli M., Meng-Papaxanthos L. Conditional generative modeling for de novo protein design with hierarchical functions // Bioinformatics. 2022. Vol. 38(13). P. 3454-3461. DOI: 10.1093/bioinformatics/btac353.
6. Yatskou M.M., Smolyakova E.V., Skakun V.V., Grinev V.V. Identification of single nucleotide genetic polymorphism sites using machine learning methods // Advances in Transdisciplinary Engineering. 2023. Vol. 42. P. 1031–1037. DOI:10.3233/ATDE231044.
7. Yatskou M.M., Smolyakova E.V., Skakun V.V., Grinev V.V. Identification of single nucleotide genetic polymorphism sites using machine learning methods // bioRxiv 2023.10.19.563060. 2023. P. 1–6. DOI: <https://doi.org/10.1101/2023.10.19.563060>.

НЕЙРОСЕТЕВОЕ ИМИТАЦИОННОЕ МОДЕЛИРОВАНИЕ ПРИ АНАЛИЗЕ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ ФЛУОРЕСЦЕНТНОЙ СПЕКТРОСКОПИИ

Н.Н. Яцков, В.В. Апанасович

*Белорусский государственный университет, Минск, Беларусь
E-mail: yatskou@bsu.by*

Рассмотрены нейросетевые подходы к имитационному моделированию экспериментальных данных при исследовании сложных биомолекулярных систем методами флуоресцентной спектроскопии. Результаты экспериментов могут быть имитационно смоделированы с помощью традиционных, физических и генеративно-состязательных нейронных сетей. Основным преимуществом нейросетевых подходов над классическим имитационным моделированием является быстрый и относительно точный способ моделирования. Нейросетевое моделирование может быть потенциально эффективным для различных приложений флуоресцентной спектроскопии, где регулярное обучение сетей не требуется.

Ключевые слова: флуоресцентная спектроскопия; имитационное моделирование; нейронная сеть.

ВВЕДЕНИЕ

Современные методы флуоресцентной спектроскопии с временным разрешением позволяют регистрировать большие наборы кинетических кривых затухания флуоресценции биофизических систем [1]. Среди существующих алгоритмов анализа и моделирования данных [2] можно выделить метод обработки больших наборов кинетических кривых затухания флуоресценции молекул с использованием алгоритмов имитационного моделирования и интеллектуального анализа данных. Его применение позволяет повысить точность оцененных параметров биофизических и оптических процессов, протекающих в исследуемых молекулярных системах [3]. В современных научных исследованиях наблюдается тенденция применения моделей глубокого обучения вместо физических моделей [4]. В связи с чем возникает вопрос о возможности использования моделей нейронных сетей в разрабатываемом подходе. Несмотря на широкое применение аппарата нейронных сетей для решения таких задач обработки данных, как классификация, кластеризация и аппроксимация интенсивностей флуоресценции, вопрос об использовании имитационного моделирования при исследовании сложных биофизических систем с помощью нейронных сетей все еще остаётся мало изученным.

В работе предлагается аналитический обзор нейросетевых подходов к имитационному моделированию экспериментальных данных при исследовании сложных биомолекулярных систем методами флуоресцентной спектроскопии.

НЕЙРОСЕТЕВЫЕ ПОДХОДЫ К ИМИТАЦИОННОМУ МОДЕЛИРОВАНИЮ ИНТЕНСИВНОСТЕЙ ФЛУОРЕСЦЕНЦИИ

В нейросетевых подходах к моделированию и анализу наборов многомерных данных применяются алгоритмы искусственных нейронных сетей и имитационного моделирования с целью воспроизведения и определения законов распределений основных характеристик объекта исследования E . Нейросетевые подходы, или искусственные нейронные сети, применяемые к решению задач математического моделирования, формально можно разделить на традиционные (аппроксимационные), физические (учитывающие физические основы законов описания процессов и систем) и генеративно-состязательные.

ТРАДИЦИОННЫЕ НЕЙРОННЫЕ СЕТИ

Традиционные нейронные сети состоят из одного или более нейронов, объединённых в слои. В многослойных сетях в рамках одного слоя данные обрабатываются параллельно, а в масштабах всей сети – последовательно, от слоя к слою. На входы нейронов передаются значения независимых переменных наблюдения объекта E_i или выходные сигналы нейронов предыдущего слоя. Веса входов нейронов регулируются в процессе обучения нейронной сети под решение конкретной задачи. В общем случае, для каждого из нейронов, скалярное произведение t векторов входных сигналов и настраиваемых весов поступает на функциональный преобразователь, характеризуемый функцией активации $\varphi(t)$, в качестве которой обычно используются ступенчатая, линейная, сигмоидальная, экспоненциальная, гиперболическая и прочие функции. Главная задача искусственной нейронной сети – формировать выходной сигнал y_i в зависимости от значений наблюдений объекта E_i , поступающих на ее вход [5].

При имитационном моделировании интенсивностей флуоресценции традиционные нейронные сети используются для генерации аппроксимирующих кривых с добавлением шума. В этом случае для набора конкретных данных подбирается наиболее подходящая нейронная сеть и обучается. Обученная нейронная сеть генерирует интенсивности флуоресценции в заданные моменты времени. К фотонным отсчетам смоделированной интенсивности добавляется гауссов шум.

ФИЗИЧЕСКИЕ НЕЙРОННЫЕ СЕТИ

Подход включает замену имитационной модели нейронной сетью, учитывающей описание физических процессов в объекте исследования. Нейронная сеть по сути представляет собой решение уравнения или системы уравнений на графах вычислений, в основе которых лежат разложение сложной функции на композицию более простых и эффективные вычисления производных по переменным простым функций [6]. Так известные библиотеки Theano и TensorFlow предназначены для автоматического дифференцирования на графах, что используется при обучении нейронных сетей. Физические (математические) модели – это системы интегро-дифференциальных уравнений или схем имитационного моделирования физических процессов. Их аналитические решения, в том числе реализованные в имитационной модели, могут быть преобразованы к графу вычислений, которому ставится в соответствие нейронная сеть. Таким образом, схему имитационного моделирования можно представить в виде графа и подобрать оптимальную нейронную сеть.

ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫЕ НЕЙРОННЫЕ СЕТИ

Идея имитационной генеративно-состязательной нейронной сети (ГНС) состоит в рассмотрении случайных сигналов некоторого скрытого пространства, представленного смоделированным набором данных, и использовании преобразующих и декодирующих этапов алгоритмов вариационных автоэнкодеров при имитации «реальных» экспериментальных сигналов [6, 7].

Имитационная модель генерирует наборы обучающих «реальных» данных в исходном временном пространстве – осуществляется моделирование интенсивностей флуоресценции с заданной физической моделью и фиксированным сочетанием ее параметров. Формируется нейронная сеть генератора, на вход которого подаются случайные величины. В качестве генератора можно использовать декодирующий слой вариационного автоэнкодера, применяемого к имитационно смоделированным данным, или специально разработанную сверточную нейронную сеть [8]. Далее используется принцип ГНС – оценки генеративных моделей с помощью состязательного процесса нейронных сетей генератора и дискриминатора. Дискриминатор – это сложный критерий качества, оценивающий схожесть смоделированного (нейросетью генератора) и реального наборов данных. Генератор обучается создавать «искусственную» гистограмму интенсивности флуоресценции, в то время как дискриминатор должен отличить искусственную гистограмму от истинной, или экспериментальной, которая может представлять собой экспериментальный набор данных или смоделированную гистограмму флуоресценции с за-

данными параметрами. Состязание нейронных сетей генератора и дискриминатора в конечном итоге приводит к созданию очень реалистичных, искусственных гистограмм. Таким образом, осуществляется обучение ГНС или подгон сгенерированных нейросетью данных под «реальные» имитационно смоделированные или экспериментальные данные. Используя хорошо обученную ГНС можно сопоставить гистограммы интенсивностей флуоресценции с физическими параметрами исследуемых биофизических систем [8].

ЗАКЛЮЧЕНИЕ

Представлен аналитический обзор нейросетевых подходов к имитационному моделированию экспериментальных данных флуоресцентной спектроскопии при исследовании сложных биомолекулярных систем. Основным преимуществом нейросетевых подходов над классическим имитационным моделированием является быстрый и относительно точный способ моделирования, недостатком – высокие вычислительные затраты в обучении сетей. Нейросетевое моделирование может быть потенциально эффективным для приложений флуоресцентной спектроскопии, где регулярное обучение сетей не требуется.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Yatskou M. M., Apanasovich V. V.* Simulation modelling and data mining approach for the study of applied fluorescence spectroscopy systems // *Journal of the Belarusian State University. Physics.* 2024. Vol. 1. P. 4–15.
2. *Datta R., Gillette A., Stefely M., Skala M. C.* Recent innovations in fluorescence lifetime imaging microscopy for biology and medicine // *Journal of Biomedical Optics.* 2021. Vol. 26(7):070603. DOI: 10.1117/1.JBO.26.7.070603.
3. *Yatskou M.M., Skakun V.V., Apanasovich V.V.* Method for processing fluorescence decay kinetic curves using data mining algorithms // *Journal of Applied Spectroscopy.* 2020. Vol. 87(2). P. 333–344. DOI: 10.1007/s10812-020-01004-3.
4. *Yatskou M.M., Apanasovich V. V.* Simulation modelling and machine learning platform for processing fluorescence spectroscopy data // *Communications in computer and information science.* 2022. Vol. 1562. P. 178–190. DOI: 10.1007/978-3-030-98883-8_13.
5. *Яцков Н.Н.* Интеллектуальный анализ данных: пособие / Минск : БГУ, 2014. С. 151.
6. *Николенко С., Кадури А., Архангельская Е.* Глубокое обучение / СПб.: Питер, 2018. С. 480.
7. *Foster D.* Generative Deep Learning : Teaching Machines to Paint, Write, Compose, and Play / 2nd ed. Oreilly & Associates Inc., 2023. P. 456.
8. Generative adversarial network enables rapid and robust fluorescence lifetime image analysis in live cells / *Chen Y.I., Chang Y.J., Liao S.C. [et al.]* // *Communications Biology.* 2022. Vol. 11, N5(1):18. DOI: 10.1038/s42003-021-02938-w.

СЕКЦИЯ «БИОИНФОРМАТИКА»

ВЕБ-ИНСТРУМЕНТ ДЛЯ ЛОКАЛЬНОГО ПОИСКА САЙТОВ СВЯЗЫВАНИЯ РЕГУЛЯТОРОВ ТРАНСКРИПЦИИ В ПОСЛЕДОВАТЕЛЬНОСТЯХ БАКТЕРИАЛЬНЫХ ГЕНОМОВ

А.В. Дигрис, П.В. Вычик, Е.И. Дувалов,
В.В. Скакун, Е.А. Николайчик

Белорусский государственный университет, Минск, Беларусь,
E-mail: digris@bsu.by

Представлен веб-инструмент, предназначенный для локального поиска регуляторных последовательностей в бактериальных геномах. Работа инструмента базируется на применении моделей регуляторных мотивов из базы данных BacRegDB, ключевой особенностью которых является использование CR-тегов, контролирующих корректность применения моделей к анализу конкретного генома. Новый инструмент существенно дополняет доступные на веб-портале BacRegDB программные средства анализа регуляторной информации, содержащейся в геномных последовательностях. Работа инструмента продемонстрирована на примере анализа генома *Pectobacterium versatile* 3-2.

Ключевые слова: регуляция транскрипции; регуляторные мотивы; BacRegDB; базы данных; веб-приложение.

Введение. Использование хранящихся в нуклеотидных базах данных бактериальных геномных последовательностей в практических целях, например, при конструировании штаммов-продуцентов или в фундаментальных исследованиях регуляторных процессов, серьезно затруднено отсутствием аннотации регуляторных элементов (промоторов, операторов, терминаторов и др.). Для решения этой проблемы мы разработали программное обеспечение для аннотации известных регуляторных элементов в бактериальных геномах [1], а позднее предложили подход к идентификации операторных последовательностей *de novo* в совершенно неисследованных геномах [2]. Этот подход использует концепцию ярлыков из критичных, т. е. непосредственно контактирующих с азотистыми основаниями ДНК, аминокислотных остатков транскрипционных факторов (т. н. CR-тегов). Идентичность CR-тегов двух транскрипционных факторов одного семейства означает, что они имеют одинаковую специфичность распознавания ДНК и один и тот же мотив сайта связывания. Поэтому при совпадении CR-тега транскрипционного фактора исследуемого организма с CR-тегом известного транскрипционного фактора применение математической модели операторного мотива известного транскрипционного фактора к исследуемому геному будет корректным.

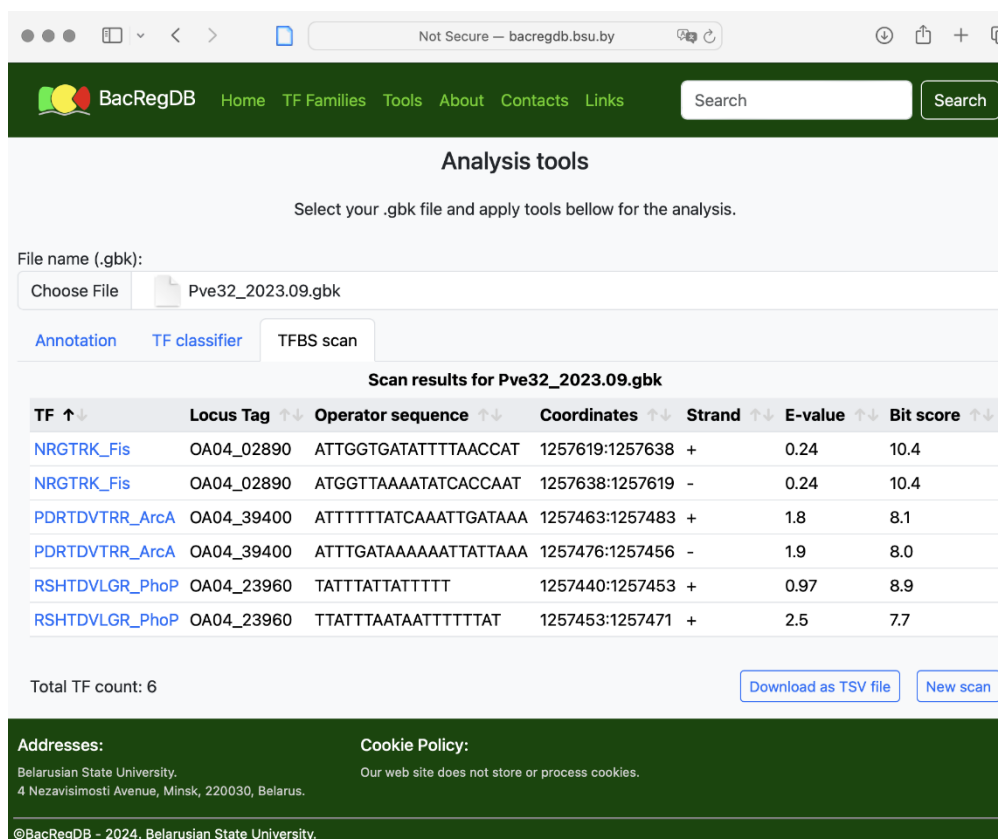
Для структурированного и целостного хранения регуляторных мотивов, ассоциированных с CR-тегами, была разработана реляционная база данных (БД) [3], а для обеспечения удобного доступа пользователей к этой базе данных организован web-портал VacRegDB. В дополнение к организации графического интерфейса доступа к базе данных, VacRegDB предоставлял пользователям два аналитических инструмента: 1) идентификатор и классификатор транскрипционных факторов и 2) средство полногеномной аннотации операторных элементов [4]. Вторым инструментом наиболее полезен для практикующих биологов и предполагает последовательное применение операторных профилей по одному к одной и той же геномной последовательности для получения максимально полной картины транскрипционной регуляции исследуемого организма. Однако чаще встречается другая практическая задача: анализ транскрипционной регуляции одного гена или оперона. В этом случае трудоемкий и времязатратный полногеномный анализ является избыточным, а более целесообразным видится организация локального поиска потенциальных транскрипционных регуляторов гена-мишени. Разработка такого инструмента – локального сканера потенциальных регуляторных мотивов, – и явилась целью настоящей работы.

Веб-инструмент. Алгоритм работы веб-инструмента «локальный сканер» основан на процедуре поиска регуляторной информации с применением программ hmmscan и nhmmscan [5] и последующей обработке их результатов с помощью скрипта, написанного на языке Python. Выбор языка Python преимущественно определялся наличием удобного функционала обработки геномных данных библиотеки Biopython [6]. Запуск локального сканера осуществляется серверной частью веб-приложения, на основе которого работает портал VacRegDB, в ответ на запрос, пришедшей от клиента. Входными данными для инструмента являются файл генома в формате .gbk и диапазон координат, в котором будет выполняться поиск. Для обеспечения приемлемого времени работы локального сканера максимальный размер регуляторной области ограничен 2000 н.п.

Локальный сканер имеет два режима работы: строгий, когда требуется отбор только тех профилей операторных мотивов, для которых CR-тег профиля соответствует множеству CR-тегов обрабатываемого генома, и нестрогий; когда такой отбор не требуется. Режим работы инструмента определяется пользователем. На первом шаге алгоритма средствами библиотеки Biopython из .gbk файла по заданным координатам экстрагируется фрагмент генома. Данный фрагмент генома в виде файла в формате FASTA подается на вход программы nhmmscan, реализующей процедуру поиска операторов. При выполнении процедуры поиска применяется биб-

лиотека из всех скрытых марковских моделей операторных мотивов, имеющих в базе данных BacRegDB [3]. При работе в строгом режиме реализуется фильтрация результатов проведенного поиска, для чего по описанному в [2] принципу формируется множество CR-тегов транскрипционных факторов, соответствующих обрабатываемому геному. Результаты работы локального сканера сохраняются в формате JSON для последующей передачи в клиентскую часть веб-приложения с целью визуализации.

Локальный сканер реализован на веб-портале BacRegDB и доступен по адресу <http://bacregdb.bsu.by/tools?toolName=scan>. Для демонстрации работы веб-инструмента выполнен поиск локальных регуляторов в геноме *Pectobacterium versatile* 3-2 в диапазоне координат 1257200-1257899 н.п. Полученные результаты представлены на рисунке.



BacRegDB Home TF Families Tools About Contacts Links Search

Analysis tools

Select your .gbk file and apply tools below for the analysis.

File name (.gbk):
Choose File Pve32_2023.09.gbk

Annotation TF classifier TFBS scan

Scan results for Pve32_2023.09.gbk

TF ↑↓	Locus Tag ↑↓	Operator sequence ↑↓	Coordinates ↑↓	Strand ↑↓	E-value ↑↓	Bit score ↑↓
NRGTRK_Fis	OA04_02890	ATTGGTGATATTTTAACCAT	1257619:1257638	+	0.24	10.4
NRGTRK_Fis	OA04_02890	ATGGTTAAATATCACCAAT	1257638:1257619	-	0.24	10.4
PDRTDVTRR_ArcA	OA04_39400	ATTTTATCAAATTGATAAA	1257463:1257483	+	1.8	8.1
PDRTDVTRR_ArcA	OA04_39400	ATTTGATAAAAAATTATATAA	1257476:1257456	-	1.9	8.0
RSHTDVLGR_PhoP	OA04_23960	TATTATATTTTT	1257440:1257453	+	0.97	8.9
RSHTDVLGR_PhoP	OA04_23960	TTATTTAATAATTTTTAT	1257453:1257471	+	2.5	7.7

Total TF count: 6 Download as TSV file New scan

Addresses:
Belarusian State University,
4 Nezavisimosti Avenue, Minsk, 220030, Belarus.

Cookie Policy:
Our web site does not store or process cookies.

©BacRegDB - 2024. Belarusian State University.

Результаты поиска регуляторных последовательностей для генома *Pectobacterium versatile* 3-2 в диапазоне координат 1257200-1257899 н.п.

В указанном интервале расположена регуляторная область дивергона *pell/pehA*, кодирующего два фактора вирулентности этой патогенной бактерии. Сканер успешно идентифицировал два сайта связывания глобального регулятора PhoP, чье присутствие в этой области было недавно продемонстри-

ровано экспериментально [7]. Также идентифицированы два сайта связывания регулятора ответа ArcA, ответственного за адаптацию к анаэробнозису, и один сайт для глобального регулятора Fis. Для Fis единственный сайт указан дважды, поскольку использован симметричный (палиндромный) профиль, а сканирование выполнено для обеих цепей ДНК.

Выводы. Разработанный веб-инструмент для локального поиска регуляторных последовательностей в бактериальных геномах, основанный на использовании CR-тегов, позволяет за короткое время определить потенциальные регуляторы конкретного бактериального гена или оперона, что очень полезно при конструировании штаммов-продуцентов и фундаментальных исследованиях транскрипционной регуляции отдельных генов/оперонов. Создание данного инструмента позволило расширить линейку средств анализа геномных данных доступных для широкого круга исследователей на веб-портале BacRegDB по адресу <http://bacregdb.bsu.by/tools>.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Nikolaichik Y., Damienikan A.U. Sigmoid: a user-friendly tool for improving bacterial genome annotation through analysis of transcription control signals. // PeerJ 4. 2016. P. e2056. DOI: <https://doi.org/10.7717/peerj.2056>
2. Nikolaichik Y., Vychik P. New approach to genome-wide automated inference of bacterial transcription factor binding sites // Abstracts of the Twelfth International Multiconference "Bioinformatics of Genome Regulation and Structure/Systems Biology (BGRS/SB-2020)" 06-10 July 2020, Novosibirsk: Institute of Cytology and Genetics. – P. 75-76. DOI: <https://doi.org/10.18699/BGRS/SB-2020-046>
3. Скакун В.В., Николайчик Е.А. Разработка базы данных мотивов регуляции транскрипции у бактерий. // Информатика. Т. 19. 2022. С. 59-71. DOI: <https://doi.org/10.37661/1816-0301-2022-19-1-59-71>
4. Дигрис А.В., Вычик П.В., Дувалов Е.И., Скакун В.В., Николайчик Е.А. BacRegDB – база данных и набор инструментов для работы с регуляторной информацией в бактериальных геномах. // Материалы 7-й Международной научно-практической конференции «Прикладные проблемы оптики, информатики, радиофизики и физики конденсированного состояния» 18-19 мая 2023 г., Минск / отв. ред. Кольчевская И.Н. 2023. С. 158-160.
5. Wheeler T.J., Eddy S.R. nhmmer: DNA homology search with profile HMMs. // Bioinformatics. Vol. 29. 2013. P. 2487-2489. DOI: <https://doi.org/10.1093/bioinformatics/btt403>
6. Biopython: freely available Python tools for computational molecular biology and bioinformatics. / Cock P.J. [et al.] // Bioinformatics. 2009. Vol. 25. P.1422-1423.
7. Kravchenko U., Gogoleva N., Kolubako A., Kruk A., Diubo J., Gogolev Y., Nikolaichik Y. The PhoPQ two-component system is the major regulator of cell surface properties, stress responses and plant-derived substrate utilisation during development of *Pectobacterium versatile*-host plant pathosystem. // Frontiers in Microbiology. 2021. Vol. 11. P. 3596. DOI: <https://doi.org/10.3389/fmicb.2020.621391>

A NOVEL ALGORITHM FOR PREDICTION OF DRUG RESISTANCE USING CROSS-ACCURACY COMPUTATION

Yu-Xiang Chen¹, Alexander V. Tuzikov²

1) *Belarusian State University, Minsk, Belarus*

2) *United Institute of Informatics problems of the National Academy of Sciences, Minsk, Belarus*

E-mail: 1c894424323@outlook.com, 2tuzikov@newman.bas-net.by

Genome-wide association studies (GWAS) are vital in determining disease etiology. Since most of the existing computational models for GWAS require many features, they are often too computationally expensive for clinical use. An algorithm for prediction of *Mycobacterium tuberculosis* drug resistance is proposed by using cross-accuracy computations. We do exhaustive investigation of SNP pairs followed by a greedy algorithm to improve prediction accuracy of SNP combinations. A dataset of 3178 *Mtb* genomes was tested to predict resistance to 8 drugs. The results were compared with those from TB-profiler and Mykrobe web systems. They indicate that our algorithm closely approximates the performance of the existing methods for first-line drugs, while is outperforming them in accuracy for the second-line drugs tested.

Keywords: drug-resistant tuberculosis; SNPs combination; cross-accuracy.

НОВЫЙ АЛГОРИТМ ПРЕДСКАЗАНИЯ ЛЕКАРСТВЕННОЙ УСТОЙЧИВОСТИ НА ОСНОВЕ ВЫЧИСЛЕНИЯ ТОЧНОСТИ ДЛЯ ПАР МУТАЦИЙ

Юйсян Чэнь¹, Александр Тузи́ков²

¹*Белорусский государственный университет, Минск, Беларусь*

²*Объединенный институт проблем информатики НАН Беларуси, Минск, Беларусь*

¹c894424323@outlook.com, ²tuzikov@newman.bas-net.by

Полногеномный анализ ассоциаций (ПАА) имеет важное значение для определения этиологии заболеваний. Поскольку большинство существующих вычислительных моделей для ПАА требуют вычисления многих признаков, они зачастую слишком трудоемкие для использования в клинической практике. В докладе предложен алгоритм прогнозирования лекарственной устойчивости микобактерий туберкулеза (*Mtb*), основанный на вычислении точности предсказания для всех пар мутаций. Для повышения точности предсказания на основе комбинаций мутаций используется жадный алгоритм. Набор данных из 3178 геномов *Mtb* был протестирован для предсказания устойчивости к 8 лекарствам. Полученные результаты сравнивались с результатами веб-сервисов TB-profiler и Mykrobe. Они показывают, что предложенный алгоритм незначительно уступает по точности предсказания этим сервисам для препаратов первой линии, но превосходит их по точности для протестированных препаратов второй линии.

Ключевые слова: лекарственно-устойчивый туберкулез; комбинация мутаций; перекрестная точность.

INTRODUCTION

Genome-wide association analysis (GWAS) is an algorithm that searches for correlations between single nucleotide polymorphisms (SNPs) and traits [1]. Due to the large number of SNPs, finding combinations of k SNPs among hundreds of thousands in genomes is a complex combinatorial problem. In recent years, Monte Carlo method [2], spanning tree method [3], machine learning methods [4, 5] etc., have appeared, which can improve search efficiency. Discriminant functions for calculating associations between SNP combinations and phenotypes are essential, and discriminant functions with light computations are favorable for improving search efficiency. Mutual information and conditional entropy (Shannon entropy-based methods) [6] and chi-square tests [7] are widely used to assess associations in a lightweight way.

The aim of this research is to explore a search method that is able to find SNP combinations satisfying high recognition accuracy. The methodology developed uses exhaustive investigation of SNPs interaction for all SNP pairs followed by a greedy algorithm to find combinations of SNPs associated with phenotypes.

METHODS

If an initial sequence of SNPs $\mathbf{x} = (x_1, x_2, \dots, x_p)$ contains thousands of elements, then one can compute accuracy values $a(x_i, x_j)$ (1) for combined SNPs created by all pairs (x_i, x_j) of single SNPs. It allows to use further this information for finding the most important combinations of SNPs.

Let us select q pairs (x_i, x_j) with the highest $a(x_i, x_j)$ values which contain distinct SNPs. Now we will propagate these pairs of SNPs to combinations of up to l SNPs by a greedy algorithm as follows. For every selected pair (x_i, x_j) consider combinations of three SNPs (x_i, x_j, x_r) for every x_r from the initial sequence $\mathbf{x}$ distinct from x_i and x_j . Select an arbitrary combination of three SNPs (x_i, x_j, x_k) with the maximum accuracy value $a(x_i, x_j, x_k)$. Similarly, one can get combinations of up to l SNPs starting from the selected pair (x_i, x_j) . Finally, let us select either a single SNP or a combination of SNPs associated with resistance to a considered drug with the maximum accuracy value.

The quality of prediction of a phenotype based on a SNP in a considered genome position can be evaluated using several measures:

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP+FN}, \text{Specificity} = \frac{TN}{TN+FP}, \\ \text{Precision} &= \frac{TP}{TP+FP}, \text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \end{aligned} \quad (1)$$

Prediction of drug resistance is based on the presence of a SNP under consideration in a genome sample. If a sample is resistant to drug d and is predicted correctly, then the prediction is considered true positive (TP). Similarly, if a sample is sensitive to drug d and is predicted to be susceptible, then the prediction is considered true negative (TN). However, no prediction is perfect, and if a sample is resistant but is predicted to be sensitive, then the prediction is considered false negative (FN). Similarly, if a sample is sensitive to drug d but is predicted to be resistant, then the prediction result is false positive (FP).

RESULT AND DISCUSSION

The original data set used in this study included the drug susceptibility test data (DST) and genome-wide data corresponding to these cases. These data were taken from the TB portal [8], presenting an excellent platform for investigations on drug-resistant tuberculosis (TB). The DST data provide verified information on resistance or sensitivity of *Mtb* samples to considered drugs. Our dataset contained 3178 whole genomes. We investigated resistance to 4 first-line drugs such as Isoniazid (INH), Rifampin (RIF), Streptomycin (SM), Ethambutol (EMB), and 4 second-line drugs, such as Amikacin (AMK), Linezolid (LZD), Levofloxacin (LFX) and Kanamycin (KM).

The efficiency of the developed algorithm was compared with Mykrobe [9] and TB-Profler [10]. Both are considered as practical prediction tools for *Mycobacterium tuberculosis* drug resistance. The prediction results of "Mykrobe v0.13.0" and "TB-Profler v5.0.1" were computed for the same dataset.

Our algorithms showed slightly underperformance than Mykrobe and TB-profler for first-line drugs, however, for second-line drugs we have got higher accuracy values than Mykrobe and TB-Profler. The algorithm is clearly applicable to maximize other predictors as well. It has also an advantage of providing results that are clearly interpretable, since resistance to a specific drug is predicted by the presence of the corresponding individual mutations in a genome sample under consideration.

Comparison of the developed algorithms with Mykrobe and TB-profiler systems

Drug	Recall	Specificity	Precision	Accuracy	Method
INH	90.3%	91.8%	96.1%	90.7%	Our algorithm
	89.4%	95.7%	97.9%	91.3%	Mykrobe
	89.9%	95.2%	97.7%	91.5%	TB-profiler
RIF	93.8%	80.9%	87.9%	88.6%	Our algorithm
	92.6%	84.8%	90.1%	89.5%	Mykrobe
	93.8%	84.5%	90.0%	90.1%	TB-profiler
SM	84.8%	69.9%	82.1%	79.1%	Our algorithm
	94.7%	64.7%	81.4%	83.3%	Mykrobe
	93.0%	66.2%	81.7%	82.8%	TB-profiler
EMB	91.9%	70.3%	69.5%	79.5%	Our algorithm
	86.5%	78.8%	75.0%	82.1%	Mykrobe
	93.2%	72.0%	71.0%	81.0%	TB-profiler
AMK	94.9%	58.3%	75.9%	79.5%	Our algorithm
	32.6%	95.3%	90.6%	58.9%	Mykrobe
	33.2%	95.0%	90.2%	59.1%	TB-profiler
LFX	92.7%	49.2%	73.6%	75.5%	Our algorithm
	50.5%	90.0%	88.5%	66.1%	Mykrobe
	53.5%	89.4%	88.6%	67.7%	TB-profiler
LZD	93.0%	80.1%	83.3%	86.7%	Our algorithm
	0.3%	100.0%	100.0%	48.6%	Mykrobe
	0.4%	100.0%	100.0%	48.6%	TB-profiler
KM	70.0%	80.0%	75.7%	75.3%	Our algorithm
	78.1%	68.1%	68.5%	72.8%	Mykrobe
	79.8%	66.6%	68.0%	72.8%	TB-profiler

CONCLUSION

We proposed a novel algorithm for predicting *Mycobacterium tuberculosis* drug resistance using cross-accuracy computations. The results obtained demonstrate the superiority of the developed algorithm over Mykrobe and TB-Profiler for the second-line drugs tested. They showed also that the number of SNPs in combinations can be constrained by 5 to ensure an appropriate approximation of prediction accuracy using the developed algorithm. This can also be utilized to reduce the computational efforts required to search for SNP combinations.

REFERENCES

1. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. / Hindorff L.A. et al.] // PNAS. 2009. P. 9362-9367. DOI: 10.1073/pnas.0903103106

2. A Bayesian model for detection of high-order interactions among genetic variants in genome-wide association studies. / Wang J. [et al.] // BMC Genomics. 2015. Vol. 16. P. 1-20. DOI: 10.1186/s12864-015-2217-6.
3. TEAM: efficient two-locus epistasis tests in human genome-wide association study. / Zhang X. [et al.] // Bioinformatics. 2010. P. i217-i227. DOI: 10.1093/bioinformatics/btq186.
4. TSLRF: two-stage algorithm based on least angle regression and random forest in genome-wide association studies. / Sun J. [et al.] // Scientific reports. 2019. Vol. 9. P. 18034. DOI: 10.1038/s41598-019-54519-x
5. Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies. / Mieth B. [et al.] // Scientific reports. 2016. Vol. 6. P. 36671. DOI: 10.1038/srep36671
6. ELSSI: parallel SNP–SNP interactions detection by ensemble multi-type detectors. / Wang X. [et al.] // Briefings in Bioinformatics. 2022. Vol. 23. P. bbac213. DOI: 10.1093/bib/bbac213.
7. Distributed multi-objective optimization for SNP-SNP interaction detection. / Li F. [et al.] // Methods. 2024. Vol. 221. P. 55-64. DOI: 10.1016/j.ymeth.2023.11.016
8. The TB portals: an open-access, web-based platform for global drug-resistant-tuberculosis data sharing and analysis. / Rosenthal A. [et al.] // Journal of clinical microbiology. 2017. Vol. 55. P. 3267-3282. DOI: 10.1128/jcm.01013-17
9. Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. / Hunt M. [et al.] // Wellcome open research. 2019. Vol. 4. P. 191. DOI: 10.12688/wellcomeopenres.15603.1
10. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. / Phelan J. E. [et al.] // Genome medicine. 2019. Vol. 11. P. 41. DOI: 10.1186/s13073-019-0650-x

РАЗРАБОТКА ПРОГРАММНОГО ОКРУЖЕНИЯ НА ОСНОВЕ DOCKER ДЛЯ ОБРАБОТКИ ДАННЫХ СЕКВЕНИРОВАНИЯ

**Н.В. Воронова-Барте, В.Б. Степаненко,
Е.Р. Яхницкая, С.С. Левыкина**

*УО «Белорусский государственный университет», Минск, Беларусь
E-mail: vikastepanenko97@gmail.com*

Разработан новый подход к анализу геномных данных через контейнерную технологию. Созданное программное окружение обеспечивает гибкую среду для обработки данных секвенирования в Docker-контейнерах, упрощая и автоматизируя процесс анализа для исследователей геномики и биоинформатики.

Ключевые слова: биоинформатика; контейнеризация; геномика.

ВВЕДЕНИЕ

С развитием высокопроизводительных методов секвенирования объем данных, генерируемых при исследованиях, значительно вырос, что повлекло за собой необходимость в разработке эффективных методов обработки и анализа геномных последовательностей. Docker представляет собой мощный инструмент для создания гибких и масштабируемых сред для анализа геномных данных. В данной работе представлен новый подход к анализу геномных данных с использованием конвейера, который включает в себя широкий спектр биоинформатических инструментов для генерации, обработки и анализа геномных последовательностей.

МАТЕРИАЛЫ И МЕТОДЫ

В работе использовались программные инструменты SPAdes, Bowtie, BWA, Samtools и Bcftools [1-5], которые были объединены в конвейер для дальнейшего создания Docker-контейнера. Также использовался язык bash для написания скриптов, решающих локальные задачи, такие как автоматизация процессов сборки Docker-образа и управление зависимостями.

Docker-контейнер представляет собой стандартизированный, изолированный и портативный пакет программного обеспечения, включающий в себя все необходимые для запуска приложений, включая код, среду выполнения, системные инструменты, библиотеки и настройки [6]. Созданный контейнер размещен в облачной репозитории - Docker Hub. Это позволяет любому пользователю получить доступ к контейнеру и использовать его на своей системе без необходимости внесения изменений в среду выполнения.

Инструменты были подобраны специально для изучения геномов эндосимбионтов, таких как выбранный тестовый объект *Buchnera aphidicola* (GCF_003099975.1).

РЕЗУЛЬТАТЫ

Схема работы конвейера изображена на рисунке.

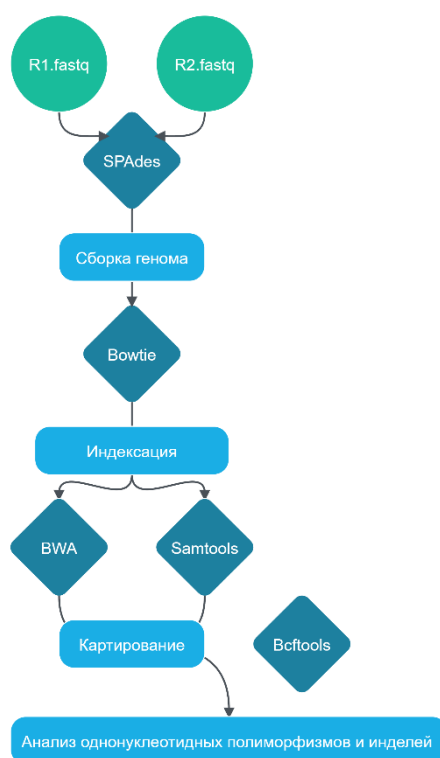


Схема работы созданного конвейера

Генерация исходного файла, содержащего собранные геномные последовательности, происходит в формате FASTA. Для этой цели используется программное обеспечение SPAdes, которое является известным в области сборки геномов и характеризуется высокой степенью эффективности [1]. Для сборки используются два файла в формате FASTQ, обозначаемые как R1 и R2. Эти файлы содержат парные прочтения, полученные из секвенирования фрагментов ДНК. Затем полученный новый файл FASTA, содержащий собранные последовательности, подвергается индексации с использованием инструмента Bowtie.

Процесс индексации позволяет эффективно представить геномные последовательности в виде структуры данных, которая оптимизирует последующее картирование геномных ридов. Такой подход способствует

ускорению и улучшению процесса анализа геномных данных, обеспечивая более быстрый доступ к информации и более точное выравнивание геномных последовательностей на референсный геном [2].

После процесса индексации следует непосредственно картирование данных на указанный референс с применением инструмента BWA (Burrows-Wheeler Aligner). Далее полученные выравнивания подвергаются обработке при помощи инструмента Samtools. Этот процесс включает в себя несколько этапов, включая сортировку выравниваний по их позиции в геноме, фильтрацию для исключения низкокачественных выравниваний или артефактов, а также индексацию выравниваний для обеспечения быстрого доступа к информации о выравниваниях в последующих анализах [7].

Обработка выравниваний с использованием инструмента Samtools создает основу для последующего анализа геномных вариантов. Точные и надежные выравнивания являются необходимым предварительным условием для корректного геномного анализа.

Последним этапом работы созданного конвейера является непосредственно проведение SNP-анализа (анализа однонуклеотидных полиморфизмов) и идентификация инделей (делеций и инсерций), что имеет важное значение для понимания генетической изменчивости и ее связи с фенотипическими характеристиками исследуемых организмов. Для этого используются инструменты Samtools и Vcftools, которые позволяют обнаруживать точные местоположения геномных вариаций, определяя отличия в последовательности нуклеотидов между исследуемыми образцами и референсным геномом [8].

Однонуклеотидные полиморфизмы и инделы могут быть ключевыми маркерами для идентификации генетических особенностей, связанных с адаптацией к окружающей среде, эволюционным процессам и различными заболеваниями.

Так, при анализе генома *Buchnera aphidicola* с использованием созданного контейнера было найдено 38 инделей, из которых: 1 индел длиной в 11 нуклеотидов (с делецией), 2 инделя длиной в 4 нуклеотида (с делецией), 3 инделя длиной в 3 нуклеотида (с делецией), 5 инделей длиной в 2 нуклеотида (с делецией), 14 инделей длиной в 1 нуклеотид (с делецией), 5 инделей длиной в 1 нуклеотид (с инсерцией), 4 инделя длиной в 2 нуклеотида (с инсерцией), 1 индел длиной в 4 нуклеотида (с инсерцией), 1 индел длиной в 6 нуклеотидов (с инсерцией), 1 индел длиной в 8 нуклеотидов (с инсерцией), 1 индел длиной в 12 нуклеотидов (с инсерцией).

Также в изучаемом геноме найдено 26 457 однонуклеотидных полиморфизмов. Обнаружено, что наиболее часто встречающиеся типы замен – А>Т, Т>А, А>Г и Т>С, в то время как наименее часто встречающиеся – С>Г, Г>С и Г>Т.

ЗАКЛЮЧЕНИЕ

Был разработан конвейер, включающий пять программных инструментов: SPAdes, Bowtie, BWA, Samtools и Bcftools. Для удобства его использования был создан Docker-контейнер, который позволяет значительно сократить время, затрачиваемое исследователями на поиск и установку соответствующих программ, выполнение отдельных геномных анализов, а также нивелирует проблему переноса настроенных приложений от одного пользователя к другому, что делает его эффективным инструментом для исследований полных геномов эндосимбионтов.

В результате в геноме *Buchnera aphidicola* было найдено 26 457 однонуклеотидных полиморфизмов. Также обнаружено, что из 38 инделей 25 были делециями длиной от 1 до 11 нуклеотидов, 13 – инсерциями от 1 до 12 нуклеотидов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Spades assembly - commandline - ABRPI-Training [Электронный ресурс]. URL: https://sepsis-omics.github.io/tutorials/modules/spades_cmdline/ (дата обращения: 23.03.2024).
2. Bowtie: Tutorial [Электронный ресурс]. URL: <https://bowtie-bio.sourceforge.net/tutorial.shtml> (дата обращения: 23.03.2024).
3. Учебное пособие по BWA-MEM [Электронный ресурс]. URL: <https://docs.tinybio.cloud/docs/bwa-mem-tutorial> (дата обращения: 23.03.2024).
4. Учебное пособие по Samtools. [Электронный ресурс]. URL: <https://docs.tinybio.cloud/docs/samtools-overview> (дата обращения: 23.03.2024).
5. BCFtools [Электронный ресурс]. URL: <https://samtools.github.io/bcftools/bcftools.html> (дата обращения: 23.03.2024).
6. Docker: Lightweight Linux Containers for Consistent Development and Deployment [Электронный ресурс]. URL: <https://www.linuxjournal.com/content/docker-lightweight-linux-containers-consistent-development-and-deployment> (дата обращения: 23.03.2024).
7. Картирование на референсный геном | Учебный сайт Александра Злобина. [Электронный ресурс]. URL: <https://kodomo.fbb.msu.ru/~alexander.zlobin/terms/third/mapping/> (дата обращения: 23.03.2024).
8. Однонуклеотидные полиморфизмы, индели и сборка | Учебный сайт Александра Злобина. [Электронный ресурс]. URL: <https://kodomo.fbb.msu.ru/~alexander.zlobin/terms/third/velvet/> (дата обращения: 23.03.2024).

ТРАНСКРИПТОМНО-ПРОТЕОМНЫЙ ПОДХОД К ИДЕНТИФИКАЦИИ НОВЫХ БЕЛКОВ ЛЕЙКОЗНЫХ КЛЕТОК ЧЕЛОВЕКА

Е.В. Гузова¹, И.Н. Ильюшёнко¹, Ф. Барне², Х.Л. Марин-Рубио³,
О. Хайденрайх², В.В. Гринев¹

1) Белорусский государственный университет, Минск, Беларусь

2) Центр детской онкологии им. Принцессы Максимы, Утрехт,
Нидерланды

3) Ньюкаслский университет, Ньюкасл, Великобритания

E-mail: k.guzowwa0099@gmail.com

В настоящей работе описан подход, позволяющий осуществлять поиск неизвестных на сегодняшний день белков, синтезирующихся в лейкозных клетках человека. В его основе лежит биоинформатический анализ данных полнотранскриптомного секвенирования указанных клеток, а также результатов протеомного анализа. Реализация такого подхода потенциально может позволить расширить репертуар неоантигенов, пригодных для противолейкозной терапии.

Ключевые слова: лейкозные клетки; транскриптом; сплайсинг РНК; протеом; неоантигены.

Среди существующих в настоящее время направлений лечения лейкозов наиболее перспективными признаны адаптивная Т-клеточная терапия и вакцинация. Главным фактором, сдерживающим их широкое применение, является низкое разнообразие молекулярных мишеней. Известно, что в роли таковых могут выступать экспрессируемые лейкозными клетками опухоль-ассоциированные или опухоль-специфические антигены. Последние, так же называемые неоантигенами, представляют наибольший интерес. Рассматриваемые главным образом как продукт экспрессии нормальных генов, мутировавших при опухолевой трансформации клеток, такие антигены характеризуются ограниченным репертуаром и зачастую недостаточной иммуногенностью.

Между тем доказано, что основным источником разнообразия транскриптома и, следовательно, протеома клеток человека является альтернативный сплайсинг РНК. Более того, многочисленные исследования показали, что его протекание, равно как и функционирование систем контроля качества РНК, в лейкозных клетках нарушено. Это, в свою очередь, позволяет сделать предположение о возможности опосредованного подобными aberrациями появления новых аминокислотных последовательностей, в том числе иммуногенных [1]. На основании изложенных фактов была поставлена цель разработать подход, который базировался бы на реконструкции полноразмерных РНК лейкозных клеток человека и позволял

с опорой на результаты протеомного анализа идентифицировать экспрессирующиеся в указанных клетках неизвестные на сегодняшний день белки. Предполагается, что, по меньшей мере, часть из них будет способна выступать в роли неоантигенов.

Обобщенная схема алгоритма поиска новых белков, экспрессирующихся в лейкозных клетках человека, представлена на рисунке ниже. В предлагаемом подходе может быть выделено два крупных блока. Первый включил в себя ряд этапов работы на уровне транскриптома исследуемых клеток и начинался с подготовки файлов, необходимых для реконструкции полноразмерных РНК. Непосредственно сборка транскриптома осуществлялась с использованием двух хорошо зарекомендовавших себя программных средств: *StringTie* и *Trinity* [4]. Первое работает с картированными прочтениями, а также эталонными аннотациями известных генов и транскриптов. Среди последовательностей РНК, реконструируемых *StringTie*, могут обнаруживаться как уже известные транскрипты, так и транскрипты, структурно не соответствующие эталонным, однако разнообразие последних ограничено. С целью нивелировать этот недостаток для выполнения рассматриваемого этапа также использовался сборщик *Trinity*, позволяющий собирать полноразмерные РНК *de novo*.



Основные этапы алгоритма поиска новых белков, экспрессирующихся в лейкозных клетках человека

Результаты работы указанных сборщиков наряду с корректно реконструированными транскриптами содержали некоторое количество шума и потому требовали проведения фильтрации. Последняя осуществлялась по

ряду параметров, выбор которых зависел от того, каким программным средством была собрана последовательность. Так, в случае реконструкции с помощью *Trinity*, фильтрация основывалась на отборе молекул в зависимости от уровня их экспрессии. Обработка результатов работы *StingTie* включала в себя учет более широкого спектра показателей. Наряду с уровнем экспрессии оценивалась протяженность транскриптов, длина экзонов и интронов в их составе, принадлежность генов той или иной хромосоме и цепь ДНК. Для установления пороговых значений количественных параметров фильтрации были построены и проанализированы соответствующие графики плотности вероятности.

Среди реконструированных транскриптов наибольший интерес представляли белок-кодирующие, поэтому на следующем этапе с использованием функционала R/Bioconductor-пакета *ORFhunteR* [3] проводилась классификация всех РНК в соответствии с наличием у них кодирующего потенциала. В транскриптах, которые были отнесены к классу кодирующих белок, устанавливались координаты основных открытых рамок считывания (ОРС). Следующий шаг заключался в трансляции *in silico* обнаруженных ОРС. Именно он обеспечивал переход ко второму блоку подхода, посвященному работе на протеомном уровне.

На следующем этапе была использована нейронная сеть *DIA-NN* [2], которая позволила провести валидацию предсказанных *in silico* белков путем поиска подтверждающих их экспрессию пептидов по данным, полученным в ходе масс-спектрометрического (МС) анализа трипсинолизованной белковой фракции лизата целевых клеток. Поскольку описываемый подход нацелен на поиск новых белков, важно было выяснить, какие из исследуемых последовательностей, чья экспрессия подтверждалась пептидами МС-анализа, несут в себе фрагменты, специфические для лейкозных клеток. В качестве таковых в данном случае рассматривались фрагменты, которые не обнаруживались в последовательностях известных на сегодняшний день белков. Последние были взяты из таких баз данных, как UniProtKB/Swiss-Prot, Consensus CDS и NCBI RefSeq.

Далее идентифицированные уникальные фрагменты, а также выравнивающиеся относительно них пептиды анализировались с использованием *blastp*. По результатам данного анализа создавался набор новых белков исследуемых клеток. Включались в него те белки, уникальные фрагменты которых, равно как и пептиды, подтверждающие их, имели идентичностью менее 100 % к известным белкам человека.

Основные результаты, полученные в ходе реализации описанного подхода с использованием в качестве исходных данных результатов полнотранскриптомного секвенирования клеток линии Kasumi-1, представлены в таблице ниже.

Основные результаты реализации подхода

	Способ сборки транскриптома	
	<i>de novo</i>	с опорой на референсный геном
Общее количество уникальных предсказанных белков	12405	25812
Количество белков, прошедших валидацию данными МС-анализа	3313	11963
Количество идентифицированных новых белков	16	11

Примечательно, что среди новых белков, предсказание которых проводилось на основе сборки полноразмерных РНК с помощью *StringTie*, 5 кодировались псевдогенами. Остальные же белки, как и следовало ожидать, были получены в результате трансляции *in silico* ОРС, идентифицированных в транскриптах, структурно не соответствовавших эталонным изоформам РНК.

Таким образом, разработан подход, который позволяет обнаруживать неизвестные в настоящее время белки лейкозных клеток человека и может рассматриваться как часть пайплайна, нацеленного на идентификацию неоантигенов указанных клеток. Его дополнение алгоритмом оценки иммуногенности обнаруживаемых новых белков позволит расширить репертуар молекулярных мишеней, ограниченность которого в настоящее время сдерживает развитие противолейкозной терапии.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. A genome-wide aberrant RNA splicing in patients with acute myeloid leukemia identifies novel potential disease markers and therapeutic targets. / Adamia S., Haibe-Kains B., Pilarski P.M. [et al.] // Clin Cancer Res. 2013. Vol. 20. P. 1135-1145. DOI: 10.1158/1078-0432.CCR-13-0956
2. DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. / Demichev V., Messner C. B., Vernardis S. I. [et al.] // Brief Communication. 2020. Vol. 17. P. 41-44. DOI: 10.1038/s41592-019-0638-x
3. Grinev V.V., Yatskou M.M., Skakun V.V., Chepeleva M.K., Nazarov P.V. ORFhunter: an accurate approach for the automatic identification and annotation of open reading frames in human mRNA molecules. // Software Impacts. 2022. № 12. P. 1-4. DOI: 10.1016/j.simpa.2022.100268
4. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. / Pertea M., Pertea G.M., Antonescu C.M. [et al.] // Nature Biotechnology. 2015. Vol. 33. P. 290-295. DOI: 10.1038/nbt.3122

ANALYSIS OF INDIVIDUAL HEALTH CHANGES BY GAIT USING A SMARTPHONE

N.V. Dorofeev, R.V. Sharapov, M.S. Goryachev

*Vladimir state university, Murom, Russian Federation
E-mail: dorofeevnnv@yandex.ru*

The work reveals the features of measuring human gait parameters using a smartphone in free mode. The results of statistical processing of accumulated data are presented: daily, weekly annual trends, results of classification of data on human diseases. The probability of correct classification is more than 0.7. It is for basic model - logarithmic regression, for such health conditions as: headache, neck pain, leg swelling, lower back pain, regulation and other pain. Based on the results of the analysis, assumptions are made about the reflection in the accelerometer data (gait parameters) of individual characteristics of the functioning of the musculoskeletal system under the influence of external and internal factors.

Keywords: health; gait; mobile phone; smartphone.

АНАЛИЗ ИНДИВИДУАЛЬНЫХ ИЗМЕНЕНИЙ В СОСТОЯНИИ ЗДОРОВЬЯ ПО ПОХОДКЕ С ПОМОЩЬЮ СМАРТФОНА

Н.В. Дорофеев, Р.В. Шарапов, М.С. Горячев

*Владимирский государственный университет имени Александра Григорьевича и Николая Григорьевича Столетовых, Муром,
Российская Федерация
E-mail: dorofeevnnv@yandex.ru*

В работе раскрываются особенности измерений параметров походки человека с применением смартфона в свободном режиме. Приводятся результаты статистической обработки накопленных данных: суточные, недельные годовые тренды, результаты классификации данных по заболеваниям человека. Вероятность верной классификации с помощью базовой модели – логарифмической регрессии, для таких состояний здоровья как: головная боль, боль в шее, отек ног, боль в пояснице, регулы и другие боли, составляет более 0.7. По результатам анализа выдвигаются предположения об отражении в данных акселерометра (параметрах походки) индивидуальных особенностей функционирования опорно-двигательного аппарата под действием внешних и внутренних факторов.

Ключевые слова: здоровье, походка, мобильный телефон, смартфон.

INTRODUCTION

Modern information technologies, data transmission systems and microelectronics make it possible to collect and analyze human biometric data in real time in the background, which is relevant in the field of personalized preventive and regenerative medicine, data protection and access control, management systems, etc. One of the developing areas is the analysis of gait

parameters using wearable and mobile devices, in particular a smartphone [1, 2]. Research on the use of smartphones to analyze gait parameters has been conducted for more than 10 years, however, they are limited to controlled laboratory conditions or the assessment of explicit relationships between gait parameters and diseases (for example, Parkinson's disease or multiple sclerosis).

The goal of this work is the development of information technologies in the field of personalized preventive medicine by identifying the relationships between human health, external and internal factors with gait parameters and developing algorithmic support for an information and analytical health diagnostic system based on wearable devices.

DATA COLLECTION AND PROCESSING

Measuring gait parameters using a mobile phone it difficult to obtain high-quality data and imposes certain restrictions on the use of a smartphone as a measuring part in everyday conditions makes [3-5]. To reduce the impact of these shortcomings on the results of the analysis, we use a data pre-processing algorithm based on correlation analysis and neural network classification of movements when forming a digital biometric profile of a person based on smartphone accelerometer data on gait parameters [5].

Long-term data collection was carried out using the proprietary AcsaActive software, available for download on the RuStore, Google Play, App Store platforms (the application is currently available in testing mode for smartphones based on iOS). Using a smartphone, volunteers recorded gait parameters while moving in a straight line over a distance of 10-15 meters (which is sufficient according to work [6]), up to three times a day, and filled out electronic questionnaires about their well-being and lifestyle.

In the resulting measurements, the segments that were most correlated with each other were selected (sections of time series from the accelerometer of a mobile phone, corresponding to a double step). The parameters of the selected segments were analyzed among themselves and it compared with the volunteer's personal data. Changes within days, weeks, and months were analyzed. Logistic regression and an ensemble of logistic regressions were used as basic models for classifying segments according to the types of diseases reflected in the questionnaire.

RESEARCH RESULTS

As a result of analyzing data on changes in the value of the cross-correlation function of segments, two types of daily trends were identified: downward and changing, when the downward trend is replaced by an upward

trend. Statistics on the manifestation of trends by day of the week are individual for each person (see Table).

An example of the distribution of types of daily trends by day of the week, in %

Trend type	Monday	Tuesday	Wednesday	Thursday	Friday
Cascading	56	61	42	58	67
Changing	44	39	58	42	33

In Fig. 1 shows the trend of changes in distance in multidimensional space between the surfaces of hyperspheres that describe points within one dimension (each point corresponds to accelerometer data for one double step), and Fig. 2 graph of changes in the measurement correlation value during the day.

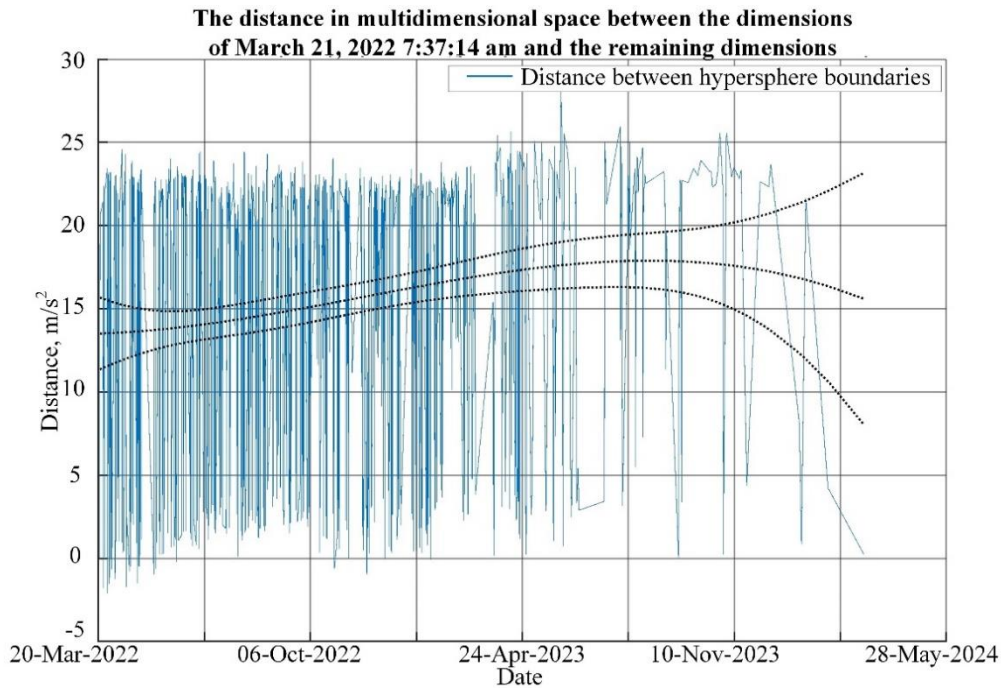


Fig. 1. Trend of changes in the distance between dimensions in multidimensional space

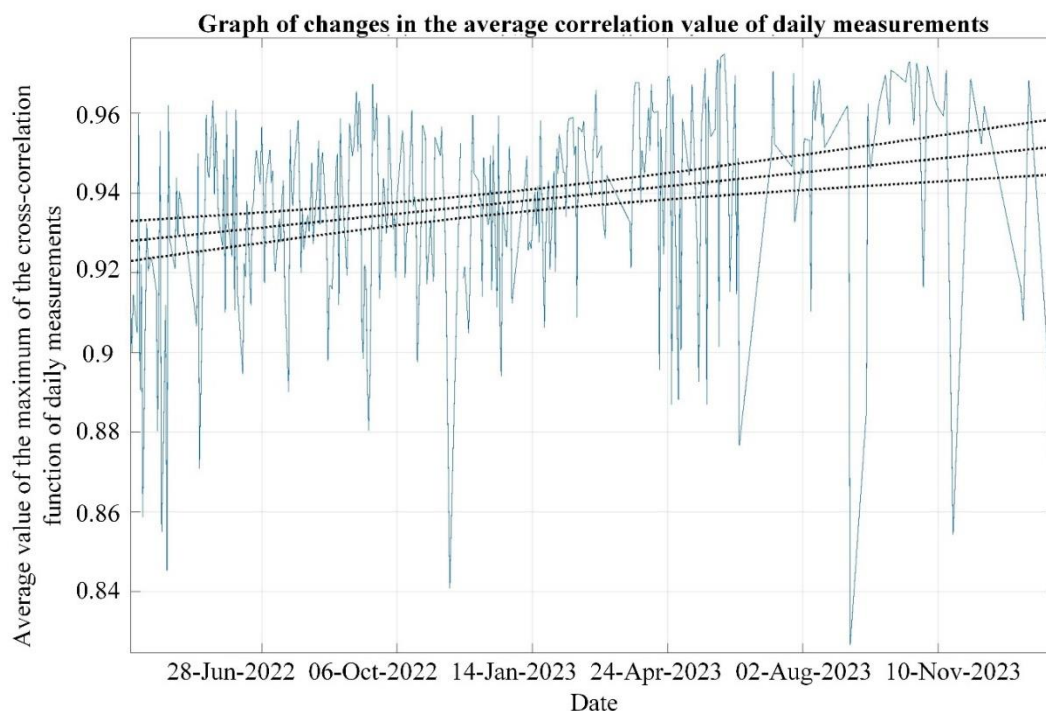


Fig. 2. Trend of change in the average correlation value of daily measurements

The use of logistic regression makes it possible to classify data segments into diseases with a probability (F-mera value) of more than 0.7. Among the diseases analyzed: headache, neck pain, swelling of the legs, lower back pain, regula and other pains.

DISCUSSION AND CONCLUSIONS

Analysis of data on a person's gait (it collected using a smartphone accelerometer in free mode over a long period) showed individual characteristics of changes in the functioning of the musculoskeletal system and the influence of external and internal factors on it. It is assumed that the features of changes in daily, weekly and monthly trends correlate with lifestyle and individual physiological characteristics. The influence of climatic factors (atmospheric pressure and ambient temperature) on a person's gait is not manifested in all volunteers, while the correlation is not higher than 0.4 when analyzing changes in the degree of similarity of segments. The high probability of classifying segments by type of disease confirms the possibility of distinguishing between events. To confirm the assumptions, further accumulation of statistical data is planned both within one volunteer and by the number of volunteers.

The research was funded by a grant from the Russian Science Foundation № 23-25-00512 <https://rscf.ru/en/project/23-25-00512>

REFERENCES

1. Human heart health prediction using GAIT parameters and machine learning model. / Singh P., Koural P.S., Mohapatra S., Kumar V. [et al.] // Biomedical Signal Processing and Control. 2024. Vol. 88. P. 1-10. DOI: <https://doi.org/10.1016/j.bspc.2023.105696>
2. *Baudendistel S.T., Haussler A.M., Rawson K.S., Earhart G.M.* Minimal clinically important differences of spatiotemporal gait variables in Parkinson disease. // Gait & Posture. 2024. Vol. 108. P. 257-263. DOI: <https://doi.org/10.1016/j.gaitpost.2023.11.016>
3. The validity of smartphone-based spatiotemporal gait measurements during walking with and without head turns: Comparison with the GAITRite[®] system. / Olsen S., Rashid U., Barbado D., Suresh P. [et al.] // Journal of Biomechanics. 2024. Vol. 162. P. 1-9. DOI: <https://doi.org/10.1016/j.jbiomech.2023.111899>
4. Validity and reliability of the Apple Health app on iPhone for measuring gait parameters in children, adults, and seniors. / Werner C., Hezel N., Dongus F., Spielmann J. [et al.] // Scientific Reports. 2023. Vol. 13. P. 1-11. DOI: <https://doi.org/10.1038/s41598-023-32550-3>
5. *Dorofeev N.V., Grecheneva A.V.* Algorithm for motion detection and gait classification based on mobile phone accelerometer data. // Izvestiya of Saratov University. Mathematics. Mechanics. Informatics. 2023. Vol. 23. P. 531-543. DOI: <https://doi.org/10.18500/1816-9791-2023-23-4-531-543>
6. *Zhou H., Zhu R., Ung A., Schatz B.* Population analysis of mortality risk: predictive models from passive monitors using motion sensors for 100,000 UK Biobank participants. // PLOS Digital Health. 2022. P. 1-22. DOI: <https://doi.org/10.1371/journal.pdig.0000045>

CLASSIFICATION OF WHOLE-SLIDE HISTOPATHOLOGICAL IMAGES OF RADIOGENIC PAPILLARY THYROID CARCINOMA

A.A. Kosareva¹, V.A. Kovalev¹, M.V. Fridman²

*The United Institute of Informatics Problems of the
National Academy of Sciences of Belarus¹, Minsk, Belarus
Belarusian State Medical University², Minsk, Belarus
E-mail: kosarevaaleksandra4317@gmail.com*

Aiming to analyse radiogenic papillary thyroid carcinoma features, different methods for preparing whole-slide histopathological images were investigated. The Bit-S R50x3 neural model was trained using four datasets: images with min-max rescaling, optical density of image values, dedicated hematoxylin and eosin channels. The best results were achieved using dataset with dedicated hematoxylin channel ($F1$ -score = 0.9504).

Key words: whole-slide images; deep learning; neural networks; data preparation.

КЛАССИФИКАЦИЯ ГИСТОПАТОЛОГИЧЕСКИХ ПОЛНО-СЛАЙДОВЫХ ИЗОБРАЖЕНИЙ РАДИОГЕННОГО ПАПИЛЛЯРНОГО РАКА ЩИТОВИДНОЙ ЖЕЛЕЗЫ

А.А. Косарева¹, В.А. Ковалев¹, М.В. Фридман²

*Объединенный институт проблем информатики Национальной академии наук
Беларуси¹, Минск, Беларусь
Белорусский государственный медицинский университет², Минск,
Беларусь
E-mail: kosarevaaleksandra4317@gmail.com*

Обсуждены методы подготовки наборов гистопатологических изображений для решения задачи классификации фрагментов папиллярной карциномы щитовидной железы. Нейросетевая модель Bit-S R50x3 была обучена на четырёх наборах данных: с использованием мин-макс. нормализации, значений оптической плотности пикселей изображений, двух отдельно выделенных каналов изображений (соответствующих позитивной реакции с ядерным красителем гематоксилином и цитоплазматическим красителем эозином). Лучший результат при обучении показал набор данных с искусственно выделенным H-каналом ($F1$ -score = 0.9504).

Ключевые слова: изображения в виде целого слайда; глубокое обучение; нейронные сети; подготовка данных.

INTRODUCTION

The vast majority of thyroid malignancies are diagnosed as papillary carcinoma [1]. As a rule, the disease does not threaten patient's life, however, clinical and histological variants of this tumour with an unfavorable prognosis are encountered [2,3]. Accordingly, to reveal specific biomarkers associated with relapse, progression or even deaths of patients are highly desirable.

“Reading” histological whole-slide images is time-consuming. A large number of characteristics can be missed during a routine investigation. On the other hand, using artificial intelligence methods in the field of whole-slide image analysis opens up new prospects for precise and reproducible detection of carcinoma features that are important for diagnosis and prognostication. Therefore, in this paper we aimed to compare several image datasets for establishing the best approach to deal with whole-slide images of papillary thyroid carcinoma. The following tasks were identified to reach objective:

1. Dataset labeling;
2. Preliminary preparation of tiles of whole-slide images;
3. Training a neural network on four datasets prepared in different ways;
4. Determining the best way to prepare images when solving the problem of binary classification of images of normal tissue and tissue with the presence of papillary carcinoma features.

This paper is organized in three sections:

1. The *Materials* section provides complete information about the prepared dataset for the development of a histopathological image analysis application. The current study used a dataset including images containing carcinoma features and images of normal tissue.

2. The *Method* section describes methods for preparing tiles of whole-slide histopathology images.

3. The *Results* section shows the results of training a neural network model on four datasets of images prepared in different ways. The metric for analysis is *F1-score*.

MATERIALS

The image dataset for the study was combined from whole-slide histopathological images of 129 patients. All these individuals were subdivided into two groups including patients born before the accident at the Chernobyl nuclear power plant (n=104, 35 men and 69 women), and their counterparts born after the disaster (n=25, 14 men and 11 women).

The materials for this study were represented by fragments of images of papillary thyroid carcinoma. The microscopic analysis was performed using routine hematoxylin and eosin staining. Histological slides containing samples of tumor and surrounding tissue were scanned with maximum resolution using the Aperio AT2 scanner system by Leica Biosystems, Germany.

Whole-slide images were labeled by an experienced histopathologist. The labeling resulted in the following characteristics: tumor size, type of histological structure, architectonics, type of tumor growth, exact localization in the thyroid, type and extension of mononuclear infiltrates, invasion of lymphatic and blood vessels,

presence of layered microcalcifications (psamoma bodies), fibrosis, and background pathology.

For this study labeled images were evenly distributed between two classes: normal tissue images and images with carcinoma features.

Finally, each image class included 1900 image tiles of whole-slide images of 256 x 256 pixels in size. The test dataset contained of around 500 image tiles in each class.

METHODS

The *Bit-S R50x3* model was used as a neural network model for the study. The choice of normalization method was based on the classification results of the trained neural network model. Classification was carried out into two classes: images with and without carcinoma features.

In the first step, the images were normalized using min-max scaling. Min-max normalization, rescaling is the simplest method and consists in rescaling the range of features to scale the range in $[0, 1]$ (fig.1, a, fig.2, a).

The following normalization steps were considered:

1. Normalizing images using Optical Density (*OD*) values. The values were determined according to the described method [4] using formulas:

$$OD = -\log_{10}(I), \quad (1)$$

$$S = V^{-1}OD, \quad (2)$$

where I – the value of the normalized color image vector in the range of $[0;1]$, S and V - saturation vectors of two-color spots corresponding to different dyes during coloring (fig.1, b, fig.2, b).

2. Using a dedicated channel corresponding to hematoxylin (*H*-channel) (fig.1, c, fig.2, c).

3. Using a dedicated channel corresponding to eosin (*E*-channel) (fig. 1, d, fig. 2, d).

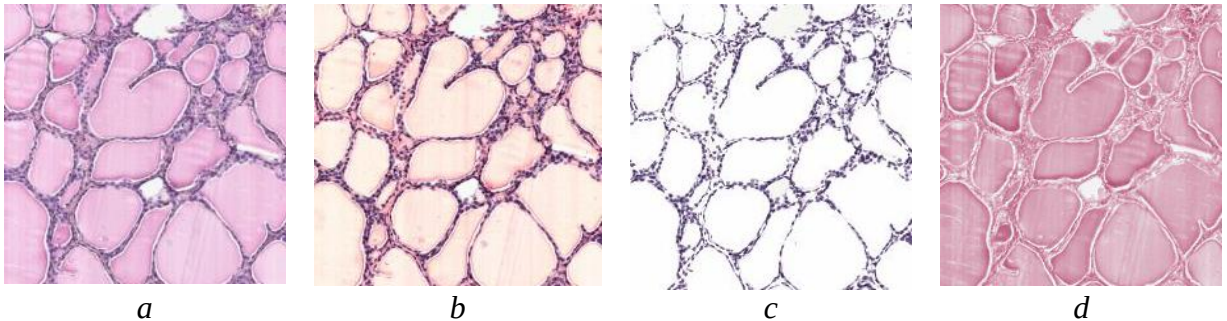


Fig. 1. Original image of normal tissue was analysed with: min-max rescaling (a), Optical Density (b), dedicated channel corresponding to hematoxylin (c), dedicated channel corresponding to eosin (d)

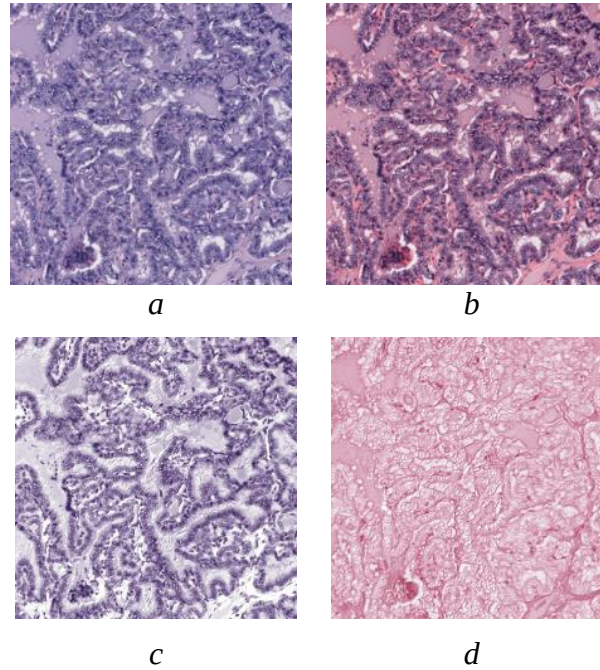


Fig. 2. Original image of papillary thyroid carcinoma was analysed with: min-max rescaling (*a*), Optical Density (*b*), dedicated channel corresponding to hematoxylin (*c*), dedicated channel corresponding to eosin (*d*)

RESULTS

F1-score [5], calculated from the training results for four test datasets, is presented in Table 1.

Result of training a Bit-S neural network model on different datasets

Image dataset	<i>F1-score</i> on the test dataset
Original images with min-max rescaling	0.8951
<i>OD</i> -images	0.9454
<i>E</i> -channel images	0.9194
<i>H</i> -channel images	0.9504

Using images with selected H-channel, it was demonstrated the best results achieves *F1-score* = 0.9504. Image optical density values generally had a positive effect on classification. Besides, separate *E*-channel showed the worst result when classifying tiles with features of carcinoma. Therefore, it can be concluded that the trained *Bit-s R50x3* model has great capacity to detect nuclei and more precisely classify tiles-containing carcinoma features.

In the future, it is planned to use the selected image preparation method for the development an application for searching and quantitative description of tiles of histopathological whole-slide images of papillary thyroid carcinoma.

REFERENCES

1. Scott A., Hundahl S.A., Fleming I.D., Fremgen A.M., Menck H.R. A national cancer data base report on 53,856 cases of thyroid carcinoma treated in the U.S., 1985-1995. // Cancer. 1998. Vol. 83. P. 2638-22948.
2. Bogolyubova A.V., Abrosimov A.Iu., Selivanova L.S., Belousov P.V. Histopathological and molecular ge-netic characteristics of clinically aggressive variants of papillary thyroid carcinoma. // Arkhiv Patologii. 2019. Vol. 81. P. 46-51.
3. Fridman M., Lam A.K., Krasko O., Schmid K.W., Branovan D.I., Demidchik Y. Morphological and clinical presentation of papillary thyroid carcinoma in children and adolescents of Belarus: the influence of radiation exposure and the source of irradiation. // Exp. Mol. Pathol. 2015. Vol. 98. P. 527-531. DOI: <https://doi.org/10.1016/j.yexmp.2015.03.039>.
4. Macenko M., Niethammer M., Marron J., Borland D., Woosley J., Guan X., Schmitt C., Thomas N. A method for normalizing histology slides for quantitative analysis. Proceedings // 2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2009. P. 9. DOI: 1107-1110. 10.1109/ISBI.2009.5193250
5. The truth of the F-measure, 2007 [Electronic resource]. URL: <https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day3/F-measure-YS-26Oct07.pdf> (date of access: 17.02.2023).

ВИРТУАЛЬНЫЙ СКРИНИНГ ПОТЕНЦИАЛЬНЫХ ИНГИБИТОРОВ ПРОНИКНОВЕНИЯ ВИЧ-1 С ПОМОЩЬЮ МЕТОДОВ МОЛЕКУЛЯРНОГО МОДЕЛИРОВАНИЯ

Я.В. Лайков¹, А.В. Тузиков¹, А.М. Андрианов²

1) Объединенный институт проблем информатики, Минск, Беларусь

2) Институт биоорганической химии НАН Беларуси, Минск, Беларусь

E-mail: laykovyan270599@gmail.com

Методами молекулярного моделирования проведен виртуальный скрининг библиотеки биологически активных молекул, направленный на идентификацию потенциальных ингибиторов домена NHR ВИЧ-1 – участка белка gp41, критически важного для слияния мембран и проникновения вируса в клетки хозяина. Для решения этой задачи использованы методы молекулярного докинга, молекулярной динамики и расчета свободной энергии образования комплексов лиганд/NHR-gp41. В результате проведенных исследований идентифицированы шесть соединений-лидеров, проявляющих высокое сродство к домену NHR белка gp41 ВИЧ-1. Предполагается, что эти соединения формируют перспективные базовые структуры для разработки эффективных и безопасных противовирусных агентов, способных ингибировать проникновение ВИЧ-1 в клетку хозяина.

Ключевые слова: ВИЧ-1; белок gp41; домен NHR; виртуальный скрининг; молекулярный докинг; молекулярная динамика; анти-ВИЧ препараты.

ВВЕДЕНИЕ

Несмотря на многие годы интенсивных исследований, на сегодняшний день ВИЧ-1 по-прежнему остается одним из наиболее серьезных инфекционных заболеваний, которое продолжает распространяться по всему миру. В свою очередь, лекарственные препараты, которые используются в клинике при высокоактивной антиретровирусной терапии, имеют ряд серьёзных недостатков, к числу которых в первую очередь следует отнести следующие: необходимость внутримышечных инъекций, высокая стоимость лекарств, возможное влияние на нормальные физиологические функции клеток, развитие приобретенной резистентности. В связи с этим, чрезвычайно важной задачей является разработка новых, эффективных и безопасных ингибиторов ВИЧ-1 широкого спектра действия. Одним из вариантов решения этой проблемы является перепрофилирование лекарственных средств, под которым понимается процесс определения новых терапевтических показаний для зарегистрированных ранее и доступных препаратов. Среди терапевтических мишеней тримерного комплекса гетеродимеров gp120/gp41, опосредующего проникновение ВИЧ-1 в клетку хозяина, важное значение имеет домен NHR белка gp41 ВИЧ-1, участвующий в образовании 6-ти спирального пучка 6-НВ, играющего ключевую

роль в процессе слияния мембран и проникновения вируса в клетку-мишень [1].

В настоящей работе с помощью стратегии перепрофилирования лекарств осуществлен компьютерный скрининг виртуальной библиотеки биологически активных соединений, направленной на идентификацию потенциальных ингибиторов домена NHR белка gp41 ВИЧ-1.

МАТЕРИАЛЫ И МЕТОДЫ

Соединения из молекулярной библиотеки, включавшей 28 806 биоактивных молекул из химических баз данных DrugBank, ZINC15 и Selleck Chemicals, были подготовлены к расчетам с использованием программных пакетов RDKit (<https://www.rdkit.org/>) и MGLTools (<https://ccsb.scripps.edu/mgltools>). Молекулярный докинг проводили с помощью программы AutoDock Vina [2] в приближении жесткого рецептора и гибких лигандов. Структуру пептида N36, имитирующего домен NHR белка gp41, заимствовали из Банка данных белков (PDB ID 5YB3). Ячейка для докинга включала гидрофобный карман домена NHR и имела следующие параметры: $\Delta X = 26 \text{ \AA}$, $\Delta Y = 18 \text{ \AA}$, $\Delta Z = 16 \text{ \AA}$ с центром при $X = 42 \text{ \AA}$, $Y = -9 \text{ \AA}$, $Z = -20 \text{ \AA}$. Параметр, характеризующий охват конформационного пространства, задавали равным 100.

После проведения молекулярного докинга соединения из молекулярной библиотеки ранжировали по значениям экспоненциального консенсусного ранга [3] на основе данных оценочных функций AutoDock Vina [2], RFScore 4 (https://github.com/oddt/rfscorevs_binary) и NNScore 2.0 (<https://github.com/durrantlab/nnscore2>).

Перед проведением МД расчетов все небелковые компоненты в структуре домена NHR белка gp41 были удалены. Молекулярную динамику комплексов лиганд/NHR-gp41 выполняли в программном пакете Amber18 с использованием силовых полей Amber ff14SB (белок) и GAFF (лиганды). Модули Antechamber и Reduce программного пакета AmberTools18 (<https://ambermd.org/doc12/Amber18.pdf>) применяли соответственно для задания парциальных зарядов атомов (модель AM1-BCC) и восстановления атомов водорода. Комплексы лиганд/NHR-gp41 помещали в октаэдрическую коробку, добавляли растворитель (модель воды TIP3PBOX 15.0) и ионы Na^+ и Cl^- до значения ионной силы, равного 0,15 М. Систему минимизировали методами наискорейшего спуска (500 шагов) и сопряженных градиентов (500 шагов), нагревали от 0 до 300 К в течение 50 пс в рамках ансамбля NVT и термостата Ланжевена, а затем уравнивали в течение 50 пс при давлении 1,0 атм (ансамбль NPT, баростат Берендсена). На заключительном шаге систему уравнивали в

течение 0,5 нс при постоянном объеме и проводили молекулярную динамику длительностью 200 нс в изобарно-изотермических условиях при температуре 300 К и давлении 1 атм.

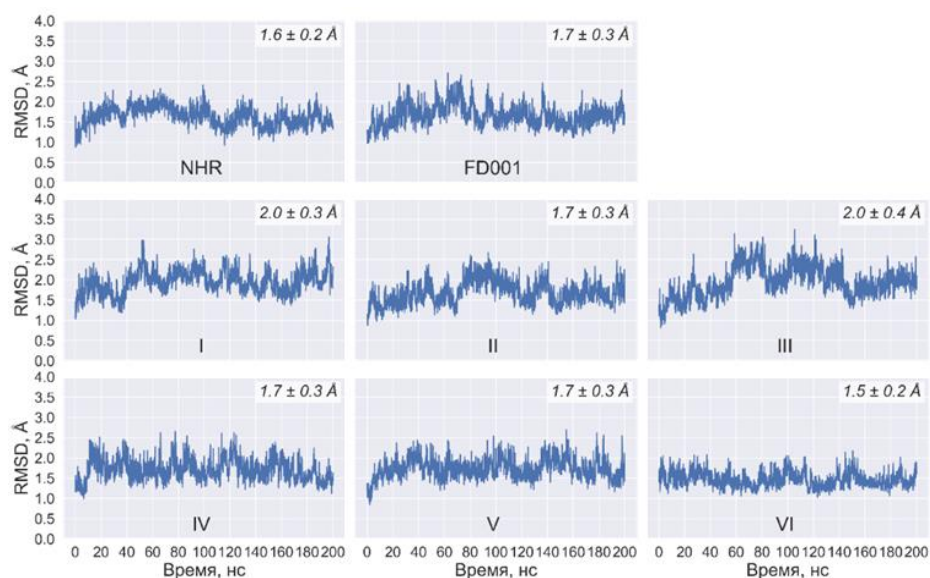
РЕЗУЛЬТАТЫ

Совместный анализ данных молекулярного докинга и МД позволил выявить 6 соединений, для которых значения свободной энергии связывания с доменом NHR белка gp41 ВИЧ-1 оказались ниже, чем полученные с использованием идентичного вычислительного протокола для контрольной молекулы FD001 (см. таблицу) – ингибитора слияния оболочечных вирусов, который блокирует образование структурного мотива 6-НВ путем взаимодействия с доменом NHR гликопротеина gp41 ВИЧ-1 [4].

Средние значения свободной энергии связывания $\langle \Delta G \rangle$, энтальпийной $\langle \Delta H \rangle$ и энтропийной $\langle T\Delta S \rangle$ составляющих свободной энергии, и соответствующие им стандартные отклонения ΔG_{STD} , $(\Delta H)_{\text{STD}}$, $(T\Delta S)_{\text{STD}}$, рассчитанные для динамических моделей комплексов идентифицированных соединений и контрольного ингибитора FD001 с NHR белка gp41 ВИЧ-1

Лиганд	$\langle \Delta H \rangle$, ккал/моль	ΔH_{STD} , ккал/ моль	$\langle T\Delta S \rangle$, ккал/ моль	$(T\Delta S)_{\text{STD}}$, ккал/ моль	$\langle \Delta G \rangle$, ккал/ моль	ΔG_{STD} , ккал/ моль
I	-37,6	6,3	-24,3	3,9	-13,2	5,1
II	-30,7	3,4	-20,1	3	-10,6	4,1
III	-32	4,5	-22,8	2,8	-9,2	4,7
IV	-30,3	4,1	-21,3	2,5	-9	4,3
V	-30,2	4,1	-21,5	2,6	-8,7	4,8
VI	-29,9	3,9	-22,1	2,9	-7,8	4,6
FD001	-30	4,2	-22,5	3,1	-7,5	4,7

Предположения о стабильности комплексов лиганд/NHR-gp41, вытекающие из данных о свободной энергии связывания (см. таблицу) подтверждают результаты анализа их МД траекторий, свидетельствующие о том, что эти комплексы не подвергаются значительным структурным перестройкам с течением времени. Этот вывод согласуется со значениями среднеквадратичных отклонений (RMSD, Root-Mean-Square Deviations) координат атомов между динамическими структурами комплексов и их исходными статическими моделями, которые близки к величинам, рассчитанным для контрольного ингибитора FD001 (рис. 1).



Временные зависимости значений RMSD (Å), рассчитанных между динамическими и стартовыми структурами комплексов лиганд/NHR-gp41. В правом верхнем углу указаны средние значения RMSD и соответствующие стандартные отклонения

ВЫВОДЫ

С помощью методов молекулярного моделирования проведен виртуальный скрининг 28 806 биологически активных молекул, направленный на идентификацию потенциальных ингибиторов ВИЧ-1, способных блокировать проникновение вируса в клетку-мишень. На основе полученных данных идентифицированы 6 соединений-лидеров, обладающих высоким сродством к домену NHR белка gp41 ВИЧ-1, что подтверждают значения свободной энергии связывания, свидетельствующие о высоком потенциале их ингибиторной активности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Atomic structure of the ectodomain from HIV-1 gp41. / Weissenhorn W., Dessen A., Harrison S.C. [et al.] // *Nature*. 1997. Vol. 387. P. 426-430. DOI: 10.1038/387426a0
2. Trott O., Olson A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. // *J. Comp. Chem.* 2010. Vol. 31. P. 455-461. DOI: 10.1002/jcc.21334
3. Exponential consensus ranking improves the outcome in docking and receptor ensemble docking. / Palacio-Rodríguez K., Lans I., Cavasotto C.N. [et al.] // *Sci. Rep.* 2019. Vol. 9. Article No. 5142. DOI: 10.1038/s41598-019-41594-3
4. The analogs of furanyl methylidene rhodanine exhibit broad-spectrum inhibitory and inactivating activities against enveloped viruses, including SARS-CoV-2 and its variants. / Pu J., He X., Xu W. [et al.] // *Viruses*. 2022. Vol. 14. P. 489. DOI: 10.3390/v14030489

ХАРАКТЕРИСТИКА UGT И GST ГЕНОВ В ГЕНОМЕ ТЛИ MACROSIPHUM ALBIFRONS

С.С. Левыкина, Д.С. Скамейко, А.Ю. Рабчун, А.М. Шульгович,
М.Д. Ракова, Н.В. Воронова-Барте

Белорусский государственный университет, Минск, Беларусь

E-mail: makarovaskam@gmail.com

Исследование суперсемейств генов UGT и GST представляет собой значимую область в изучении механизмов метаболизма ксенобиотиков у насекомых. В данной работе проводится анализ экзон-интронной структуры предсказанных UGT- и GST-генов тли *Macrosiphum albifrons*, а также сравнение полученных данных с другими видами тли.

Ключевые слова: экзон-интронная структура; тли; системы детоксикации; GST-гены; UGT-гены.

ВВЕДЕНИЕ

Macrosiphum albifrons является опасным вредителем и переносчиком инфекционных заболеваний люпина, широко используемого в качестве кормовой и декоративной культуры в Беларуси. Продолжительное питание этого вредителя приводит к снижению итогового количества и качества урожая. Для борьбы с тлей в сельском хозяйстве активно применяются различные инсектициды, однако в результате это зачастую приводит к загрязнению окружающей среды и гибели полезных насекомых-опылителей, в то время как тли способны к быстрому формированию устойчивости к применяющимся препаратам.

В рамках данного исследования было акцентировано внимание на генах UGT (уридиндифосфогликозилтрансфераза) и GST (глутатион S-трансфераза), поскольку они играют ключевую роль в системе метаболизма ксенобиотиков у насекомых [1].

МАТЕРИАЛЫ И МЕТОДЫ

Была проведена функциональная аннотация с последующим редактированием для повышения точности полученных итоговых моделей UGT- и GST-генов, которые являются консенсусными последовательностями между предсказаниями скрытых марковских моделей, белковой и РНК-гомологией, что доказывает функциональность обнаруженных генов. Курирование аннотации генных моделей проводилось с использованием геномного браузера Apollo [2] и программы BLAST [3]. Экзон-интронная структура генов была сконструирована с использованием программы GSDS 2.0 [4]. Для проведения сравнительного анализа

количества генов системы детоксикации в геноме *M. albifrons* дополнительно были извлечены данные для близкородственных видов того же таксона: *Aphis gossypii* (GCF_020184175.1), *Acyrtosiphon pisum* (GCF_005508785.1), *Melanaphis sacchari* (GCF_002803265.2). Также в работе были использованы геномные данные, полученные сотрудниками СНИЛ биоинформатики и молекулярной эволюции животных, а именно геномы *M. albifrons*, *M. rosae* и *A. craccivora*.

РЕЗУЛЬТАТЫ

В таблицах 1 и 2 представлена сравнительная характеристика GST- и UGT-генов в геномах тлей. Гены, результаты по которым не являются уникальными, исключены из таблиц, однако общее их количество указано в последней строке.

Таблица 1

Количество генов GST в геномах тлей

GST ген	<i>Macrosiphum albifrons</i>	<i>Macrosiphum rosae</i>	<i>Aphis gossypii</i>	<i>Acyrtosiphon pisum</i>	<i>Melanaphis sacchari</i>	<i>Aphis craccivora</i>
N	14	9	10	23	8	6

Таблица 2

Количество генов UGT в геномах тлей

UGT ген	<i>Macrosiphum albifrons</i>	<i>Macrosiphum rosae</i>	<i>Aphis gossypii</i>	<i>Acyrtosiphon pisum</i>	<i>Melanaphis sacchari</i>	<i>Aphis craccivora</i>
UGT5	7	0	0	0	0	0
UGT4	9	0	0	0	0	0
2A1	0	1	0	0	1	1
1-7C	0	2	1	0	2	1
1_5	0	1	1	2	1	1
342B3	1	0	0	0	0	0
...	...	...	...	...	...	...
Total	79	70	66	54	44	65

Внутри рода *Macrosiphum* было отмечено, что в геноме *M. albifrons* в семействе UGT отсутствуют гены 2A1 и 1-7C. В сравнении с другими представителями в геноме *M. albifrons* присутствуют уникальные гены UGT4, UGT5 и 342B3. Также установлено, что в геноме тли *M. albifrons* отсутствует ген 1_5, который встречается у каждого другого анализируемого вида.

Для предсказанных GST-генов установлено, что их количество в геноме *M. albifrons* уступает только количеству в геноме *A. pisum*. Экзон-интронные структуры генов UGT и GST проиллюстрированы на рис. 1 и 2.

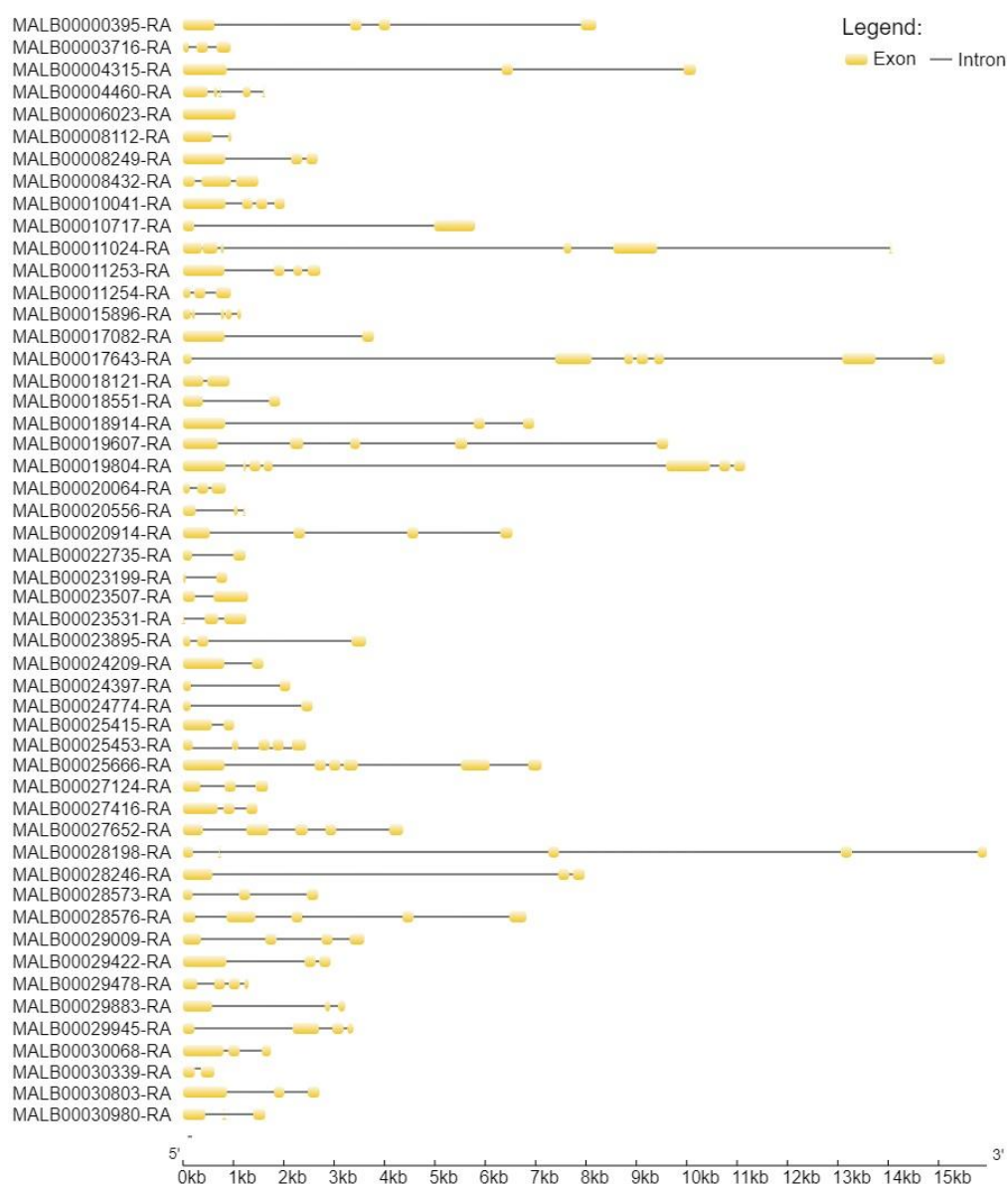


Рис. 1. Часть результатов по анализу экзон-интронной структуры генов UGT в геноме *Macrosiphum albifrons*

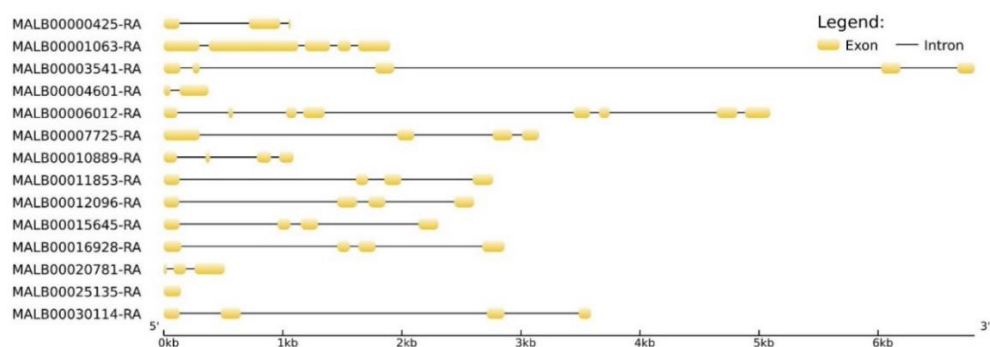


Рис. 2. Экзон-интронная структура генов GST в геноме *Macrosiphum albifrons*

Было установлено, что число интронов и экзонов сильно различается среди предсказанных UGT-генов в геноме тли *M. albifrons*. В среднем гены содержат от 1 до 4 экзонов, максимальное количество – 10. Средняя длина последовательностей составляет 3,3 кб (1 кб = 1000 нуклеотидов). Количество интронов – от 0 до 6.

В GST-генах наиболее часто встречается по 4 экзона, максимальное количество – 8, минимальное – 1; количество интронов – от 0 до 7. Последовательности имеют среднюю длину 2,5 кб.

Экзон-интронную структуру генов GST характеризуют как консервативную. Ранее в исследованиях в геномах тлей была обнаружена схожая экзон-интронная структура, выделенная в отдельное подсемейство Sigma [5]. Среди последовательностей *M. albifrons* было обнаружено, что последовательности с ID MALB00012096, MALB00015645, MALB00016929 и MALB00011853 имеют структуру, характерную для этого подсемейства, что было дополнительно подтверждено с использованием BLAST.

ЗАКЛЮЧЕНИЕ

В геноме *M. albifrons* идентифицировано суммарно 79 UGT и 14 GST предсказанных генов, охарактеризованы их экзон-интронные структуры. Анализ экзон-интронной структуры позволил идентифицировать часть GST-генов до уровня подсемейства. Установлено, что UGT-гены более вариабельны, поскольку схожих паттернов внутри генома *M. albifrons* и в геномах других видов тлей не обнаружено. Различия в генном составе между видами могут быть связаны с функциональной дифференциацией генов на протяжении эволюционного пути генома и формированием у тлей адаптаций к окружающей среде и различным кормовым специализациям.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. The role of UDP-glycosyltransferases in xenobiotic resistance. / Dimunová D. [et al.] // Drug Metabolism Reviews. 2022. Vol. 54. P. 282-298. DOI: 10.1080/03602532.2022.2083632
2. Apollo: democratizing genome annotation. / Dunn N.A. [et al.] // PLoS Computational Biology. 2019. Vol. 15. P. e1006790. DOI: 10.1371/journal.pcbi.1006790
3. BLAST+: architecture and applications. / Camacho C. [et al.] // BMC Bioinformatics. 2009. Vol. 10. P. 1-9. DOI: 10.1186/1471-2105-10-421
4. GSDS 2.0: an upgraded gene feature visualization server. / Hu B. [et al.] // Bioinformatics. 2015. Vol. 31. P. 1296-1297. DOI: 10.1093/bioinformatics/btu817
5. He H., Crabbe M.J.C., Ren Z. Detoxification gene families at the genome-wide level of rhus gall aphid *Schlechtendalia chinensis*. // Genes. 2022. Vol. 13. P. 1627. DOI: 10.3390/genes13091627

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОРТОГРУПП ВНУТРИ КЛАССА НАСЕКОМЫЕ

С.С. Левыкина, М.Д. Ракова, Н.В. Воронова-Барте

*Белорусский государственный университет, Минск, Беларусь
E-mail: rkvmry@gmail.com*

В данной работе содержатся актуальные данные по числу ортогрупп в следующих отрядах: Hemiptera, Hymenoptera, Diptera, Lepidoptera. Основной акцент работы направлен на изучение и анализ ортогрупп внутри семейства Aphidoidea (Настоящие тли) в виду их быстрого темпа развития адаптаций к ксенобиотикам и инсектицидам.

Ключевые слова: насекомые; тли; общие и уникальные ортогруппы.

ВВЕДЕНИЕ

Насекомые являются самой многочисленной группой живых организмов на Земле, играя ключевую роль в экосистемах и обладая высоким уровнем биологического разнообразия.

Тли – насекомые фитофаги, которых к настоящему времени известно более 5200 видов. Многие из них являются опасными вредителями растений, имеющих для человека сельскохозяйственную ценность [1]. Наличие у тлей множественных механизмов резистентности предполагает, что устойчивость может возникнуть к любому классу инсектицидов при наличии достаточного времени, постоянного давления отбора и достаточно большой популяции для проведения отбора [2].

Ортогруппа – это набор генов, произошедших от гена последнего общего предка всех рассматриваемых видов. Изучение ортогрупп проводится на ядерных геномах с целью получения кластеров по всем генам, что позволяет далее выделить и работать с кластерами, содержащими гены-интереса, например, гены системы детоксикации ксенобиотиков, изучение которых позволит разработать качественные меры по контролю численности насекомых.

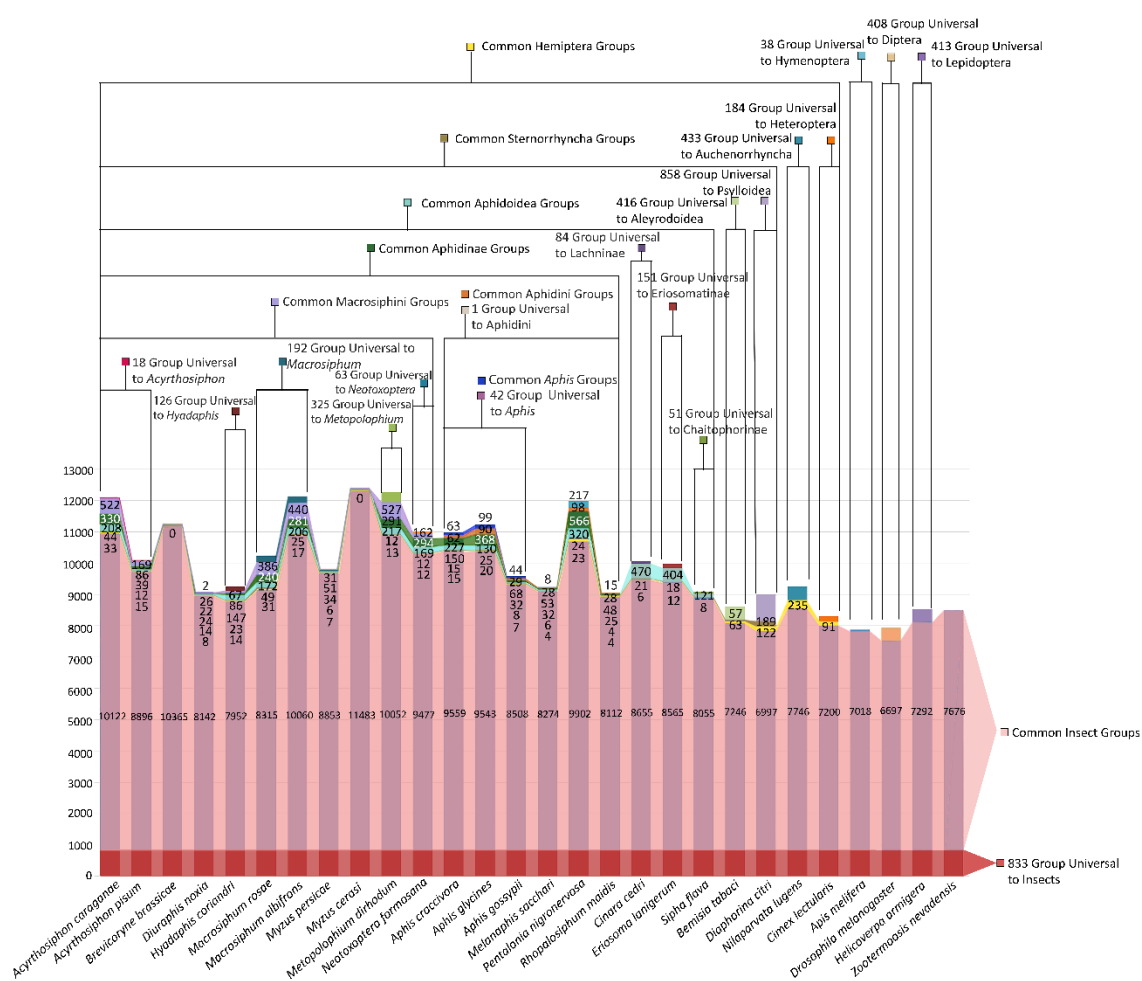
МАТЕРИАЛЫ И МЕТОДЫ

Был проведен анализ белковых последовательностей полных геномов представителей класса Insecta (Насекомые), извлечение которых было проведено из базы данных RefSeq (15 геномов), а также белковых последовательностей из полных геномов тлей, полученных в СНИЛ биоинформатики и молекулярной эволюции животных (13 геномов). Таким образом, выборка была сформирована суммарно из 28-ми ядерных геномов насекомых. С помощью скрипта, написанного с использованием команд-

ной оболочки Bash, была проведена процедура очистки скачанных последовательностей от изоформ. Поиск ортогрупп был проведен в программе OrthoFinder [3].

РЕЗУЛЬТАТЫ

Распределение уникальных и общих групп ортологов, найденных в результате проведенного анализа, отображено на графике, представленном ниже.



Результаты анализа ортогрупп для представителей класса Insecta

За уникальные группы принимали те, которые могли встречаться не у всех представителей таксона, но в других таксонах однозначно не встречались. За общие – последовательности, которые встречались у всех представителей таксона, но в других таксонах отсутствовали. Было установлено, что число уникальных ортогрупп для класса Insecta составляет 833,

а общее число групп ортологов варьируется в пределах от 6697 у *D. melanogaster* (Diptera) до 11483 у *M. cerasi* (Hemiptera), в среднем это значение составляет 8612 ортогрупп на вид.

На уровне отрядов уникальные ортогруппы в количестве 38, 408 и 413 были найдены в геномах Hymenoptera, Diptera, Lepidoptera соответственно, в то время как в отряде Hemiptera уникальных ортогрупп найдено не было. В свою очередь, отряд Hemiptera включает в себя несколько подотрядов: Sternorrhyncha (Грудохоботные), Auchenorrhyncha (Цикадовые), Heteroptera (Клопы); в последних двух подотрядах были найдены уникальные ортогруппы в количестве 433 и 184 соответственно. Наибольшее количество ортогрупп в подотряде Sternorrhyncha у *D. citri* (Hemiptera) – 189, а наименьшее число общих ортогрупп выявлено у *B. brassicae* (Hemiptera) и *M. cerasi* (Hemiptera) – 0. В целом, наибольшее число общих ортогрупп внутри отряда Hemiptera обнаружено в геноме *N. lugens* (подотряд Auchenorrhyncha) – 235 группы.

В семействе Aphidoidea, принадлежащего к подотряду Sternorrhyncha, наибольшим числом общих ортогрупп (470) обладает *C. cedri*, уникальных ортогрупп выявлено не было. Уникальные ортогруппы были найдены в семействах Aleyrodoidea (Белокрылки) в геноме *B. tabaci* – 416 ортогрупп и в семействе Psylloidea (Листоблошковые) в геноме *D. citri* – 858 ортогрупп. В подсемействе Aphidinae наибольшее количество общих ортогрупп найдено в геномах *P. nigronervosa* (566) и *A. glycines* (368), уникальные ортогруппы обнаружены не были. В других подсемействах присутствуют уникальные ортогруппы для Lachninae – 84 (*C. cedri*), в подсемействе Eriosomatinae – 151 (*E. lanigerum*), Chaitophorinae – 51 (*S. flava*).

В таблице ниже отдельно представлена численная характеристика уникальных ортогрупп для представителей триб подсемейства Aphidinae. Результаты показывают, что наибольшим числом уникальных и общих групп ортологов в трибе Macrosiphini обладает вид *M. dirhodum* – 325 и 527 ортогрупп соответственно. В трибе Aphidini наибольшее количество как уникальных, так и общих ортогрупп найдено у представителя *P. Nigronervosa* – 217 и 98, меньше всего уникальных и общих ортогрупп обнаружено в геноме *Melanaphis sacchari* – 8 и 28 соответственно. Аналогичное число общих ортогрупп отмечено также в геноме *R. maidis*. Таким образом, можно сделать вывод о том, что триба Macrosiphini является более вариативной и насыщенной по содержанию общих и уникальных ортогрупп, чем триба Aphidini.

ЗАКЛЮЧЕНИЕ

По результатам проведенного анализа, можно сделать вывод о том, что наибольшее количество уникальных ортогрупп было обнаружено в отрядах Diptera (408) и Lepidoptera (413), что может указывать на генетическое и функциональное разнообразие генов у представителей этих таксонов. Отряд Hemiptera выделяется отсутствием уникальных ортогрупп, что может быть связано с консервативностью генов внутри этого отряда. Однако внутри подсемейства Aphidinae представители трибы Macrosiphini имеют более высокий уровень генетической изменчивости и адаптивности, чем представители трибы Aphidini.

Количество уникальных ортогрупп у представителей подсемейства Aphidinae

Триба	Вид	Код доступа в RefSeq	Число уникальных групп генов	Число общих групп генов
Macrosiphini	<i>Acyrtosiphon caraganae</i>	Не депонирована	18	522
	<i>Acyrtosiphon pisum</i>	GCF_005508785.1	18	169
	<i>Brevicoryne brassicae</i>	Не депонирована	0	0
	<i>Diuraphis noxia</i>	GCF_001186385.1	2	26
	<i>Hyadaphis coriandri</i>	Не депонирована	126	67
	<i>Macrosiphum rosae</i>	Не депонирована	192	386
	<i>Macrosiphum albifrons</i>	Не депонирована	192	440
	<i>Myzus persicae</i>	GCF_001856785.1	0	31
	<i>Myzus cerasi</i>	Не депонирована	0	0
	<i>Metopolophium dirhodum</i>	Не депонирована	325	527
	<i>Neotoxoptera formosana</i>	Не депонирована	63	162
Aphidini	<i>Aphis craccivora</i>	Не депонирована	42	62
	<i>Aphis glycines</i>	Не депонирована	42	90
	<i>Aphis gossypii</i>	GCF_020184175.1	42	29
	<i>Melanaphis sacchari</i>	GCF_002803265.2	8	28
	<i>Pentalonia nigronervosa</i>	Не депонирована	217	98
	<i>Rhopalosiphum maidis</i>	GCF_003676215.2	15	28

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Vilcinskas A. Biology and ecology of aphids. // Germany: CRC Press, 2016. P. 272.
2. Nauen R., Elbert A. European monitoring of resistance to insecticides in *Myzus persicae* and *Aphis gossypii* (Hemiptera: Aphididae) with special reference to imidacloprid. // Bull Entomol Res. 2003. Vol. 93. P. 47-54. DOI: 10.1079/BER2002215
3. Emms D.M., Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. // Genome Biology. 2015. Vol. 16. P. 1-17.

ИДЕНТИФИКАЦИЯ СТРУКТУРНЫХ ВАРИАЦИЙ В ГЕНОМЕ КЛЕТОК ОСТРОГО МИЕЛОИДНОГО ЛЕЙКОЗА ПО ДАННЫМ CAPTURE Hi-C

М.А. Лиховец, Е.В. Гузова, Н.Н. Яцков, В.В. Скакун, В.В. Гринев

Белорусский государственный университет, Минск, Беларусь
E-mail: likhovetsmax@gmail.com

Структурные вариации генома могут быть важными драйверными мутациями онкогенеза. В то же время их обнаружение по-прежнему является сложной задачей. В настоящей работе представлены результаты поиска структурных вариаций в геноме клеток модельной линии острого миелоидного лейкоза человека Kasumi-1 по данным Capture Hi-C. На основании этих данных были получены координаты потенциальных перестроек в геноме изучаемых клеток. Для части обнаруженных вариаций получены подтверждения их транскрипционной активности.

Ключевые слова: Capture Hi-C; 3D структура хроматина; структурные вариации; транскрипция; лейкозные клетки человека.

ВВЕДЕНИЕ

В последние десятилетия наблюдается активное развитие методов определения 3D структуры хроматина с использованием секвенирования нового поколения. Наиболее популярными и одними из самых мощных среди них являются методы, основанные на технологии высокопроизводительного захвата конформации хромосом, или Hi-C. Данная технология позволяет получить частоты всех попарных контактов между участками хроматина в ядре. Как правило, изначально такие данные получают для изучения особенностей архитектуры генома и обнаружения различных регуляторных взаимодействий. Однако было показано, что данные Hi-C можно использовать для таких задач, как сборка генома или обнаружение протяженных мутаций – структурных вариаций. Структурные вариации включают делеции, инверсии, дупликации и транслокации. Они могут приводить к изменению копийности генов, менять регуляцию их экспрессии, а также приводить к образованию гибридных (онко)генов [1].

ОБНАРУЖЕНИЕ СТРУКТУРНЫХ ВАРИАЦИЙ

В настоящей работе использовались данные Capture Hi-C для клеток модельной линии острого миелоидного лейкоза человека Kasumi-1. Этот метод представляет собой модификацию Hi-C, при которой библиотека для секвенирования обогащается фрагментами, содержащими промоторы интересующих генов (всего более 20000 различных промоторов) [2].

В дальнейшем при обработке данных использовались координаты захваченных при процедуре Capture Hi-C фрагментов, которые были разработаны

по эталонному геному hg19. В связи с этим изначально данные Capture Hi-C картировались на геном hg19, а обработанные результаты были переведены в координаты hg38 при помощи HiCLift [3]. Парноконцевые прочтения были получены из SRA с кодами доступа SRR7521620 и SRR7521621. Предварительная обработка, картирование на эталонный геном, фильтрация артефактов и дубликатов полимеразной цепной реакции была проведена при помощи пайплайна HiCUP [4]. Из полученных пар взаимодействующих фрагментов с помощью Python пакета cooler рассчитывалась матрица контактов с разрешением 10Kb [5]. Учитывая особенности процедуры Capture Hi-C, была также проведена нормализация матрицы контактов, что необходимо для надежного предсказания структурных вариаций. Кроме того, картированные и обработанные прочтения были использованы для определения существенных взаимодействий между участками генома, успешно проведенного благодаря функционалу R/Bioconductor пакета CHiCAGO [6].

Матрицы контактов Hi-C могут быть визуализированы в виде тепловых карт, где частоты взаимодействия между двумя участками хроматина отображаются интенсивностью окраски соответствующего пикселя. Для идентификации различных структурных вариаций была использована сверточная нейронная сеть EagleC, способная классифицировать паттерны на картах Hi-C, соответствующие вариациям структуры [7]. Ключевым результатом работы EagleC является список геномных координат мест разрыва-соединения хромосом(-ы), или так называемых брейкпоинтов (BP, от англ. breakpoint) и их ориентация.

В клетках Kasumi-1 с помощью EagleC было идентифицировано множество BPs, которые в дальнейшем были отфильтрованы против событий, попадающих в области центромер, а также событий, имеющих низкую подтверждаемость. Всего после фильтрации осталось 16 BPs, координаты части из которых представлены в таблице. Эти координаты соответствуют эталонному геному hg38 и имеют разрешение +/- 10Kb.

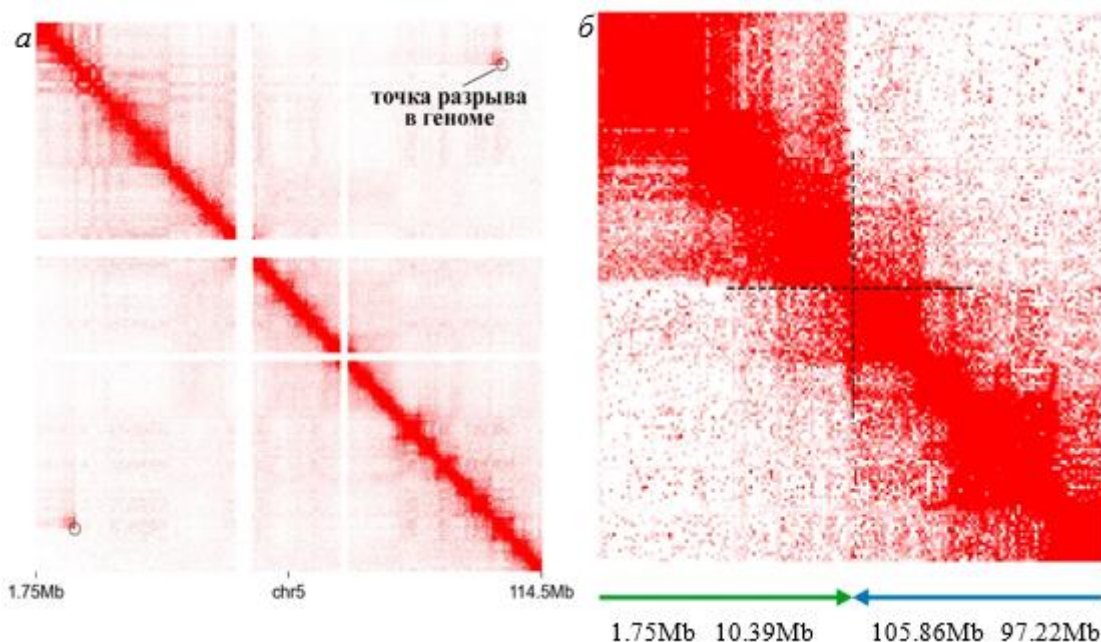
Полученные координаты точек разрыва соотносятся с опубликованными ранее данными о структурных вариациях в клетках Kasumi-1 [8]. Кроме того, достоверность полученных предсказаний подтверждается обнаружением известной для этих клеток реципрокной транслокации t(8;21)(q22;q22). Однако ряд BPs, обнаруженных в представленной работе, не был ранее описан.

В завершение нами была предпринята попытка выяснить ведут ли обнаруженные нами геномные перестройки к появлению новых транскрипционно активных гибридных (онко)генов.

Координаты наиболее значимых ВР в геноме клеток Kasumi-1

Хромо- сома ВР № 1	Координата ВР № 1	Хромо- сома ВР № 2	Координата ВР № 2	Вероятность ориентации			
				3'-3'	3'-5'	5'-3'	5'-5'
15	46377802	9	29300002	4,6e-19	3,7e-24	9,4e-28	1,0
2	203465277	8	122677761	1,1e-15	1,0	7,5e-11	2,7e-13
21	34857703	8	92047772	1,9e-20	1,0	0,9	7,0e-19
16	16776143	16	18086143	4,0e-20	1,7e-16	1,0	6,5e-19
5	10389888	5	105894299	1,0	1,1e-24	1,5e-15	2,1e-14
3	113211153	3	114471153	3,9e-15	3,5e-10	1,0	4,2e-14

Для этого мы сопоставили геномные координаты ВР, идентифицированных по данным Capture Hi-C, с координатами ВР, которые были обнаружены нами ранее по данным полнотранскриптомного секвенирования RNA-Seq. Учитывая разрешение матриц контактов и потенциальное отличие координат ВР из RNA-seq из-за сплайсинга, мы провели поиск ВР из RNA-Seq, координаты которых находятся в пределах +/- 50Kb от соответствующего ВР из данных Capture Hi-C. В результате был найден транскрипт гибридного онкогена RUNX1-RUNX1T1, соответствующий известной реципрокной транслокации t(8;21)(q22;q22), а также транскрипт, соответствующий ВР на хромосоме 5. Данный ВР ранее не был описан, его координаты соответствуют гену MARCHF6 и межгенной области (рисунок).



Карты контактов Hi-C для участка хромосомы 5 без учета структурной вариации (а) и с учетом (б)

ЗАКЛЮЧЕНИЕ

Таким образом, на основании данных о пространственной организации хроматина были получены координаты потенциальных ВРs в геноме клеток Kasumi-1. Одна из вновь обнаруженных точек рекомбинации, согласно данным транскриптомного секвенирования, приводит к образованию гибридного (онко)гена, состоящего из двух удалённых участков генома. Тем не менее, представленные в данной работе результаты требуют последующей валидации экспериментальными методами.

Особый интерес представляет более подробное изучение изменения структуры хроматина в областях соответствующих перестроек и изучение влияния такой структуры на экспрессию расположенных рядом генов. Кроме того, полученные данные являются основой для дальнейшего поиска ассоциаций между гибридными транскриптами, обнаруживаемыми по данным RNA-Seq, и 3D организацией хроматина.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Genome-wide detection of enhancer-hijacking events from chromatin interaction data in rearranged genomes. / Wang X., Xu J., Zhang B. [et al.] // Nat. Methods. 2021. Vol. 18, № 6. P. 661-668. DOI: 10.1038/s41592-021-01164-w
2. RUNX1-ETO depletion in t(8;21) AML leads to C/EBP α - and AP-1-mediated alterations in enhancer-promoter interaction. / Ptasinska A., Pickin A., Assi S.A. [et al.] // Cell Rep. 2019. Vol. 28, № 12. P. 3022-3031.e7. DOI: 10.1016/j.celrep.2019.08.040
3. Wang X., Yue F. HiCLift: a fast and efficient tool for converting chromatin interaction data between genome assemblies. // Bioinformatics. 2023. Vol. 39, № 6. P. btad389. DOI: 10.1093/bioinformatics/btad389
4. Wingett S.W., Ewels P., Furlan-Magaril M. HiCUP: pipeline for mapping and processing Hi-C data. // F1000Research. 2015. Vol. 4. P. 1310. DOI: 10.12688/f1000research.7334.1
5. Abdennur N., Mirny L.A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. // Bioinformatics. 2020. Vol. 36, № 1. P. 311-316. DOI: 10.1093/bioinformatics/btz540
6. CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. / Cairns J., Freire-Pritchett P., Wingett S.W. [et al.] // Genome Biol. 2016. Vol. 17, № 1. P. 127. DOI: 10.1186/s13059-016-0992-2
7. Wang X., Luan Y., Yue F. EagleC: A deep-learning framework for detecting a full range of structural variations from bulk and single-cell contact maps. // Sci. Adv. 2022. Vol. 8, № 24. P. eabn9215. DOI: 10.1126/sciadv.abn9215
8. Subtype-specific 3D genome alteration in acute myeloid leukaemia. / Xu J., Song F., Lyu H. [et al.] // Nature. 2022. Vol. 611, № 7935. P. 387-398. DOI: 10.1038/s41586-022-05365-x

РАЗРАБОТКА ПРОГРАММНОГО ПАКЕТА ДЛЯ ПОИСКА СМЕЖНЫХ КОНТИГОВ ПРИ СБОРКЕ ГЕНОМОВ DE NOVO

Н.Р. Мозоль, Л.Н. Валентович

*Белорусский государственный университет, Минск, Беларусь
E-mail: nazarmozol@mail.ru*

В данном исследовании представлена разработка программного обеспечения, направленного на автоматизацию процесса поиска и анализа смежных контигов в геномных сборках. Сборка геномов de novo метод конструирования геномов из большого количества (коротких или длинных) фрагментов ДНК без априорного знания о правильной последовательности или порядке этих фрагментов. Разработанный программный пакет реализован на языке C++, обладает высокой производительностью и эффективностью, работая быстрее по сравнению с аналогичными решениями. Это обеспечивает оперативное выполнение задач по поиску и анализу смежных контигов в геномных сборках, что значительно экономит время и ресурсы исследователей.

Ключевые слова: сборка генома; контиги; de novo сборка; биоинформатика.

Введение. Задачей сборки генома является восстановление последовательности ДНК (ее длина составляет от миллионов до миллиардов нуклеотидов у разных живых существ) на основании информации, полученной в результате секвенирования. [1].

De novo сборка не требует наличия референсного генома. Учитывая, что большинство организмов на данный момент еще не отсеквенированы, de novo сборка таких организмов может быть использована как первый этап в их изучении. Полезно собирать геном de novo даже при наличии референсного генома, так как при этом можно обнаружить участки, последовательности которых отсутствуют в предшествующей геномной сборке. Однако сборка de novo является алгоритмически сложным и вычислительно затратным процессом. Также данный подход отличается высокой чувствительностью к ошибкам.

Целью работы является разработка алгоритма и программы для поиска смежных контигов при сборке геномов de novo.

Возможности пакета. Разработанный пакет обнаруживает соседние контиги путём сопоставления их концов. Рассчитывается коэффициент LQ, отражающий степень связи между контигами:

$$LQ = \frac{n_a}{n} \cdot 100, \quad (1)$$

где n_a — это количество смежных концов контигов, а n — это общее количество концов контигов (то есть количество контигов = 2).

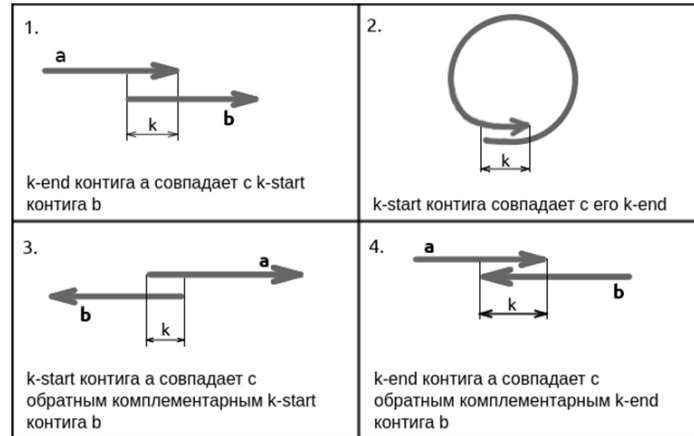


Рис. 1. Варианты смежных контигов

Ожидаемая длина генома вычисляется следующим образом:

$$L_g = \sum_{i=1}^{N_{cont}} L_i^{cont} m_i - \sum_{ovl \in O} L^{ovl}, \quad (2)$$

где L_g – это ожидаемая длина генома, N_{cont} – это количество контигов, L_i^{cont} – это длина i -го контига, O – это набор перекрытий между началами/концами контигов, L^{ovl} – это длина перекрытия ovl , m_i – это множественность i -го контига.

$$m_i = \frac{cov_i}{cov_1}, \quad (3)$$

где cov_i – это покрытие i -го контига.

Если в заголовках последовательностей нет информации о покрытии, множественность рассчитывается следующим образом:

$$m_i = \max(1.0, \min(N_S, N_E)), \quad (4)$$

где N_S – это количество контигов, которые перекрываются с началом i -го контига, N_E – с концом [4].

Сравнение с предшествующими программными пакетами. Combinator-FQ, реализованный на языке Python, основан на аналогичном алгоритме. Для сравнения был выбран единый Fasta файл, длиной минимального перекрытия были выбраны 5 пар оснований, длина максимального – 50 пар оснований. Использованы Python 3.11.4, C++ 1.19.9 [2, 3].

Зависимость времени работы программ от количества прочтений имеет характер, показанный на рис. 2. В среднем, разработанный программный пакет оказался на 20% быстрее предшественника, что делает его более выгодным в использовании.

Заключение. В ходе исследования был разработан программный пакет, направленный на автоматизацию процесса поиска и анализа смежных

контигов в геномных сборках. Этот пакет позволяет обнаруживать соседние контиги путем сопоставления их концов, что является важным этапом в сборке геномов *de novo*.

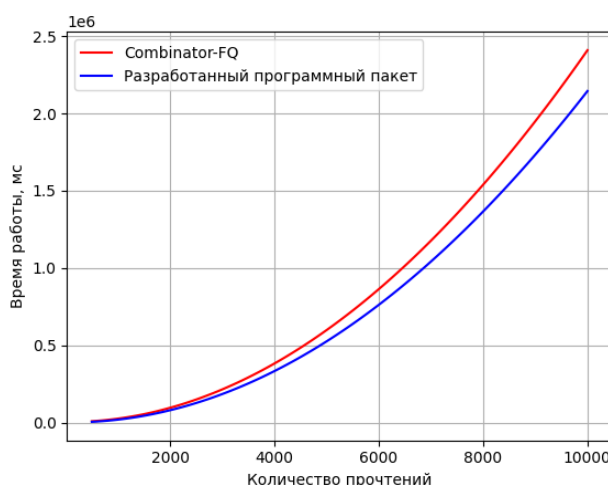


Рис. 2. График скорости выполнения программ

Помимо этого, программный пакет предоставляет разнообразную информацию о смежных контигах, включая коэффициент LQ, который отражает степень связи между контигами, а также ожидаемую длину генома. Эти данные позволяют получать дополнительные инсайты и информацию о структуре и организации генома. Программный пакет реализован на языке C++ и обладает высокой производительностью и эффективностью, что позволяет ему работать быстрее по сравнению с аналогичными решениями. Полученные результаты показывают, что разработанный программный пакет может экономить время и ресурсы исследователей при работе с геномными данными, что делает его ценным инструментом в области биоинформатики и геномики [4].

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. А.В. Александров, С.В. Казаков, С.В. Мельников, А.А. Сергушичев. Метод *de novo* сборки контигов геномных последовательностей на основе совместного применения графов Де Брюина и графов перекрытий [Электронный ресурс]. URL: https://is.ifmo.ru/works/2012/telematika/alexandrov_telematika.pdf (дата обращения: 30.03.2024).
2. Combinator-FQ [Электронный ресурс]. URL: <https://github.com/masikol/combinator-FQ/tree/main> (дата обращения: 30.03.2024).
3. Detector-Of-Adjacent-Contigs [Электронный ресурс]. URL: <https://github.com/Nazarmmm/Detector-Of-Adjacent-Contigs> (дата обращения: 30.03.2024).
4. Сиколенко М. А., Сергеев Р. С., Валентович Л. Н. Метод оценки полноты нуклеотидных данных для сборки геномных последовательностей на основе расчёта доли смежных контигов [Электронный ресурс]. URL: <https://elib.bsu.by/handle/123456789/248660> (дата обращения: 30.03.2024).

ИМИТАЦИОННАЯ МОДЕЛЬ ГЕНЕРАЦИИ САЙТОВ ОДНОНУКЛЕОТИДНОГО ПОЛИМОРФИЗМА В МОЛЕКУЛАХ ДНК ЧЕЛОВЕКА

Д. Д. Сарнацкий, Н. Н. Яцков, В. В. Гринев

Белорусский государственный университет, Минск, Беларусь
E-mail: sarnacki@bsu.by;

В работе предложена имитационная модель генерации сайтов однонуклеотидного полиморфизма в молекулах ДНК человека для обработки экспериментальных данных геномного секвенирования. В модели имитируются числа нуклеотидных прочтений в четырех каналах системы регистрации, учтены статистические линейные взаимосвязи между каналами и реализовано добавление гауссового шума для тестирования и сравнительного анализа алгоритмов определения сайтов полиморфизма.

Ключевые слова: однонуклеотидный полиморфизм; имитационное моделирование.

ВВЕДЕНИЕ

В живых организмах состав геномов варьирует. Различия называются генетическим полиморфизмом и одной из наиболее частых форм такой вариации является однонуклеотидный полиморфизм (SNP, от англ. Single Nucleotide Polymorphism). Генетические полиморфизмы влияют на фенотипы и заболевания [1]. Имитационное моделирование сайтов SNP позволяет предсказывать и изучать влияние генетических вариаций на фенотипы, заболевания и ответы на лекарственные препараты. Использование статистических оценок экспериментальных распределений в имитационных моделях генетических вариаций позволяет моделировать данные, близкие к экспериментальным [2]. В работе [2] представлена имитационная модель, основанная на бета-распределении. Основным ограничением модели является отсутствие воспроизведения числа покрытий в четырех каналах системы регистрации прочтений нуклеотидных сайтов.

В связи с этим, целью настоящей работы стала разработка имитационной модели для воспроизведения числа покрытий в четырех каналах регистрации нуклеотидных прочтений сайтов с учетом статистических взаимосвязей между каналами.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

Предположим, что сайт j содержит референсное нуклеотидное основание r (нуклеотиды А, С, G или Т); $D = \{R, N_1, N_2, N_3\}$ – набор из n прочтений (ридов) нуклеотидных оснований А, С, G или Т, покрывающих сайт j , получаемых при геномном секвенировании.

Идея алгоритма имитационного моделирования заключается в случайной генерации N_{SNP} позиций SNP-сайтов в последовательности рассматриваемой молекулы S , состоящей из N нуклеотидных сайтов, для каждого из которых сумма чисел покрытий сайта n , покрытий в референсном канале R и трех нереференсных N_1, N_2, N_3 воспроизводятся с учетом корреляционных взаимосвязей между каналами в соответствии с заданным законом распределения (формулы (1)-(4)).

Бета-распределение

$$n_b(x, \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) + \Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad (1)$$

где β и α ($\beta, \alpha > 0$) – произвольные фиксированные параметры, задающие форму кривой распределения; Γ – гамма-функция.

Нормальное распределение

$$n_g(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad (2)$$

где μ и σ – параметры, математическое ожидание и среднеквадратическое отклонение.

Распределение Вейбулла

$$n_w(x, k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}, \quad (3)$$

где k и λ ($k, \lambda > 0$) – параметры, задающие форму кривой распределения.

Гамма-распределение

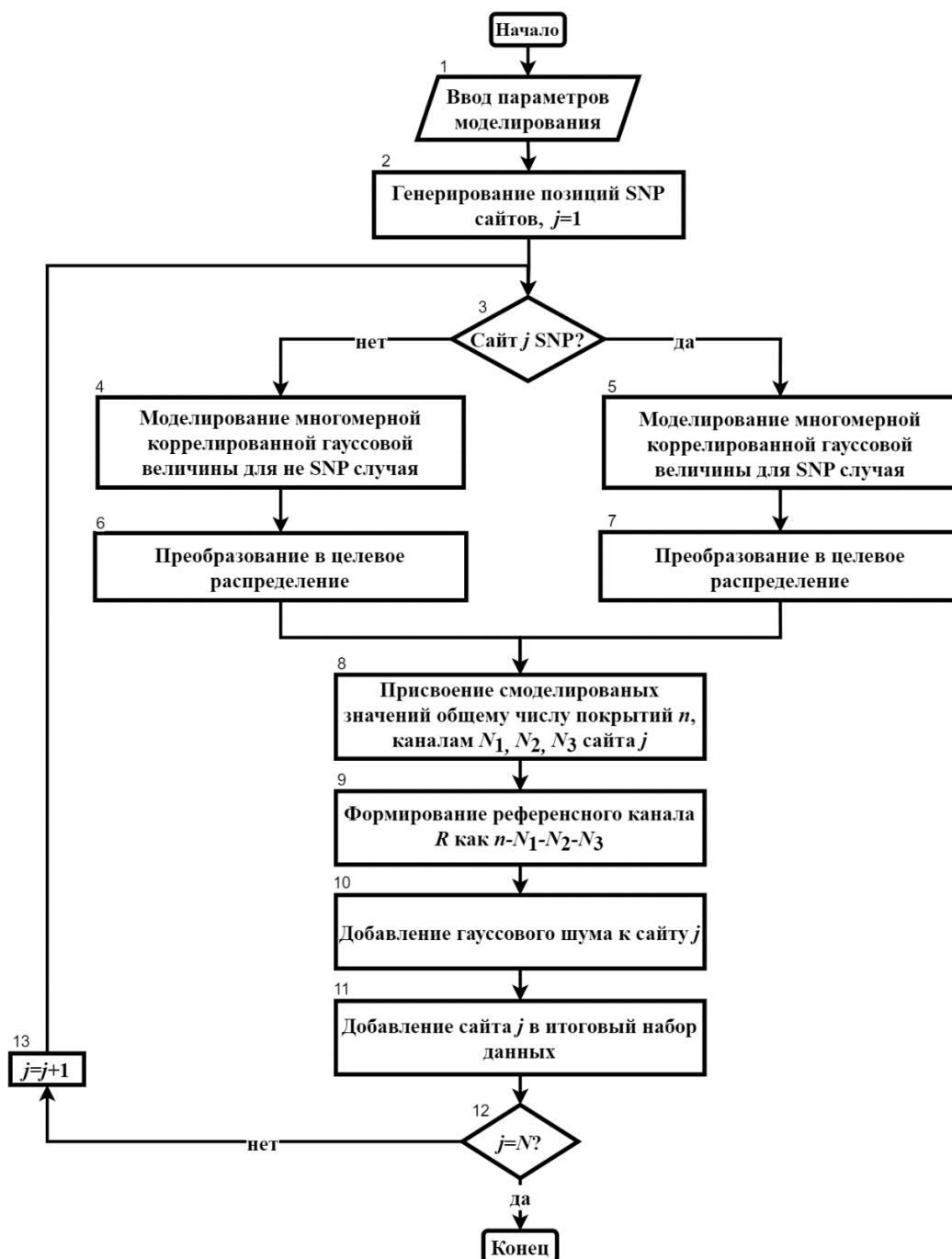
$$n_\Gamma(x, k, \theta) = x^{k-1} \frac{\exp(-\frac{x}{\theta})}{\Gamma(k)\theta^k}, \quad (4)$$

где k и θ ($k, \theta > 0$) – параметры, задающие форму кривой распределения.

Для экспериментальных данных характерно преобладание нулевых значений в нереференсных каналах. Предложено использовать при моделировании вероятности нулевых значений p_1 для канала N_1 и условные вероятности $p_2(0|N_1)$ и $p_3(0|N_2)$ для каналов N_2 и N_3 . Если $\sum_{i=1}^3 N_i \geq n$, то $R = 0$.

Для моделирования линейно-зависимых гауссовых случайных величин используется метод Халецкого с последующим преобразованием к заданному распределению. Реализовано добавление гауссового шума для тестирования и сравнительного анализа алгоритмов определения сайтов однонуклеотидного полиморфизма.

Блок-схема алгоритма имитационного моделирования однонуклеотидных полиморфизмов представлена на рисунке.



Блок-схема алгоритма моделирования

РЕЗУЛЬТАТЫ

Рассмотрены экспериментальные данные 22-ой хромосомы генома HG001 [2]. Удалены нуклеотидные сайты с числом покрытий меньше 10. Гистограммы чисел покрытий n , R , N_1 , N_2 , N_3 , построенные по экспериментальным данным, аппроксимированы функциями плотностей нормального, бета-, Вейбулла и гамма-распределений. Наилучшие результаты получены

для гамма-распределения. Полученные экспериментальные оценки параметров распределений используются в имитационном моделировании. В качестве проверки адекватности разработанной модели рассмотрены диаграммы соответствия смоделированных и экспериментально верифицированных плотностей распределений для чисел покрытий n , N_1 , N_2 , N_3 и точность восстановления моделируемых параметров при оценке параметров распределений.

Параметры имитационных моделей попадают в 95% доверительные интервалы оценок параметров, что позволяет предположить об адекватности имитационных моделей, а именно – о корректности процедур моделирования числа покрытий сайтов в четырех каналах по нормальному, бета, Вейбулла и гамма законам распределений.

Для исследования алгоритмов идентификации сайтов SNP смоделированы данные с учетом добавления гауссового шума с параметром σ . Варьирование параметром σ изменяет уровень экспериментального шума, а именно – регулирует информативность полезного сигнала, что позволяет всесторонне исследовать эффективность разрабатываемых или выбранных алгоритмов идентификации сайтов SNP и воссоздавать специальные экспериментальные условия.

ЗАКЛЮЧЕНИЕ

Разработана имитационная модель для воспроизведения чисел покрытий в четырех каналах регистрации нуклеотидных прочтений сайтов с учетом статистических взаимосвязей между каналами. Наивысшая точность моделирования числа покрытий установлена для Гамма распределения. Предложенная имитационная модель учитывает статистические линейные взаимосвязи между каналами регистрации числа прочтений, что позволяет более точно исследовать характеристики нуклеотидных сайтов. При сравнительном анализе методов обнаружения сайтов однонуклеотидных полиморфизмов добавление гауссового шума позволяет учесть инструментальную ошибку секвенирующего устройства и повысить точность имитационного моделирования.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Sung W.K. Algorithms for next-generation sequencing. // Chapman & Hall/CRC Computational Biology Series. 2017. P. 175-180.
2. Yatskou M.M., Smolyakova E.V., Skakun V.V., Grinev V.V. Simulation modelling for machine learning identification of single nucleotide polymorphisms in human genomes. // Pattern Recognition and Information Processing (PRIP'2023). Artificial Universe: New Horisont: Proceedings of the 16th International Conference, October 17–19, 2023, Belarus, Minsk / Belarusian State University: eds. A. Nedzved. A. Belotserkovsky. Minsk: BSU, 2023. P. 49-53.

ДИЗАЙН И АНАЛИЗ СТРУКТУРЫ ГИБРИДНОГО БЕЛКА ANGPT1_VEGF165 ДЛЯ ЛЕЧЕНИЯ ИШЕМИИ НИЖНИХ КОНЕЧНОСТЕЙ

А.Б. Саченко, Е.А. Попичева, В.В. Щур, А.В. Янцевич

*Институт биоорганической химии НАН Беларуси, Минск, Беларусь
E-mail: antosuk@yandex.ru*

В данной работе проведен анализ траекторий молекулярной динамики и пространственных структур гомологичных моделей гибридных белков ANGPT1_VEGF165 с различными линкерами. Установлено, что более предпочтительным являются гибкие линкеры GGGsGGGGsGGGsGGGs, GGGGsGGGGsGGGGs, GGGsGGGGs для ANGPT1_VEGF165 гибридного белка и PAPAPAP, GGGGsGGGGs, GGGsGGGGsGGGGsGGGs, GGGGsGGGGsGGGGsGGGGs для VEGF165_ANGPT1 гибридного белка.

Ключевые слова: молекулярная динамика; гибридный белок; ANGPT1; VEGF165; ангиогенез.

Введение. Хроническая ишемия нижних конечностей является одной из самых распространенных и актуальных патологий сердечно-сосудистой системы человека. Среди населения старше 50 лет распространенность этого заболевания составляет 5-8%, а при наличии таких факторов риска, как курение, гиперлипидемия, артериальная гипертензия или сахарный диабет достигает 30% [1].

Ишемия конечностей чаще всего вызвана атеросклерозом и окклюзией в периферических артериях. Прогрессирование хронической ишемии нижних конечностей приводит к ампутации с той же частотой, что и несколько десятков лет назад, а традиционные методы лечения заболевания показывают недостаточно высокую эффективность. Поэтому существует явная необходимость в новых эффективных методах лечения пациентов, страдающих от ишемии конечностей во всем мире [2]. В настоящее время перспективным методом лечения ишемии нижних конечностей является терапевтический ангиогенез, основанный на введении в ишемизированные ткани генетических конструкций с белками активаторами ангиогенеза VEGF165 (изоформа эндотелиального фактора роста сосудов длиной в 165 аминокислот) и ANGPT1.

Препараты на основе плазмидной ДНК с генами VEGF165 и ANGPT1 продемонстрировали высокую эффективность при лечении хронической ишемии ПА и ПБ степени [3]. Комбинированное применение препаратов на основе генов VEGF165 и ANGPT1 позволяет с большей эффективностью формировать новые сосуды для устойчивой неоваскуляризации, чем применение только одного из препаратов. Создание генетической конструкции кодирующей гибридный белок ANGPT1_VEGF165

может существенно увеличить терапевтический эффект и расширить спектр действия препарата [2].

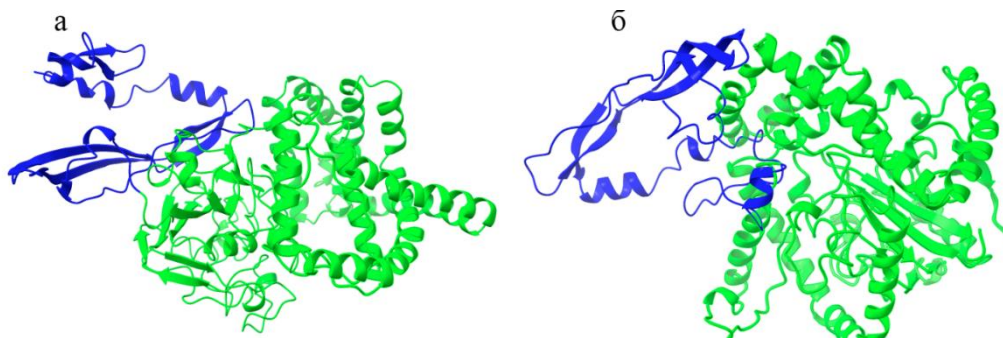
Для определения оптимального расположения доменов, а также установления оптимальной длины и аминокислотного состава пептидного линкера между двумя белками проведен анализ траекторий молекулярной динамики и пространственных структур гомологичных моделей гибридных белков ANGPT1_VEGF165.

Методы. Модели гибридных белков построены методом гомологичного моделирования (комбинацией онлайн сервиса Robetta [4] и программы Modeller 10.0). В качестве шаблонов для ANGPT1 и VEGF165 использовали гомологичные модели белков, представленные в базе данных Protein Data Bank (PDB): 1DEQ, 1EI3, 1LWU, 1M1J, 1Z3U, 3GHD, 4EPU, 4JYO, 4JZC, 4ZFG, 1FZV, 1MJV, 1MKG, 1RV6, 1WQ8, 1WQ9, 3P9W, 3V2A, 3V6B, 4DEQ, 5FV1, 5O4E, 6T9D, 6ZFI. Оптимизированные для экспрессии в клетках человека аминокислотные последовательности ANGPT1 и VEGF165 взяты из экспрессионных векторов, полученных нами ранее, а использованные линкеры выбраны из литературных источников, где они применялись для создания гибридных белков.

Валидацию полученных моделей проводили с помощью программы ChimeraX v 1.6 (визуальная проверка моделей на правильное формирование функциональных доменов и фолдинга белка), а также при помощи онлайн сервиса SWISS-Model Structure Assessment [5] (по карте распределения двугранных углов и показателю QMEAN – описывающему основные геометрические аспекты белковой структуры).

Дальнейшую валидацию моделей проводили с использованием метода молекулярной динамики (МД) в программе AMBER16 в силовом поле Amber (набор параметров ff14). Перед МД систему подвергали процедуре оптимизации: минимизация в течение 20000 шагов с использованием метода сопряженных градиентов, нагрев в течение 5 нс до температуры 310 К (NVT ансамбль). Моделирование свободной динамики проводили в течение 50 нс (NPT ансамбль) при температуре в 310 К и постоянном давлении в 1 атм.

Результаты. Присоединение белка VEGF165 по N концу белка ANGPT1 требует линкера длиной не менее 10 аминокислот. Это обусловлено нахождением N-конца белка ANGPT1 внутри белковой глобулы (рис. 1), что существенно усложняет её доступность. С конец белка ANGPT1 более доступен и позволяет использовать линкеры длиной от 5 аминокислот. В результате первичного осмотра моделей выбраны 66 моделей (из 95) гибридных белков для дальнейшей их проверки методом МД.



Модели гибридных белков с присоединением VEGF165 (а, синий) по С концу ANGPT1 (зеленый) и с присоединением (б) VEGF165 по N концу ANGPT1

В результате анализа МД пространственных структур полученных моделей определены (с наилучшими показателями стабильности, незначительные колебания СКО, радиуса гирации и изменения количества образованных и разрушенных водородных связей, а также наименьшим радиусом гирации) оптимальные гибкие линкеры GGGsGGGGsGGGsGGGs, GGGGsGGGGsGGGGs, GGGsGGGGs для ANGPT1_VEGF165 гибридного белка и PAPAPAP, GGGGsGGGGGs, GGGsGGGGsGGGGsGGGs, GGGGsGGGGsGGGGsGGGGs для VEGF165_ANGPT1 гибридного белка. Выбранные линкеры обеспечивают высокую стабильность гибридного белка и подвижность доменов относительно друг друга.

Вывод. В ходе анализа траекторий МД и пространственных структур гомологичных моделей гибридных белков VEGF165_ANGPT1 и ANGPT1_VEGF165 определены предпочтительные линкеры, которые будут использованы для создания соответствующих генетических конструкций, кодирующих гибридные белки для лечения ишемии нижних конечностей.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. ACC/AHA guidelines for the management of patients with acute myocardial infarction: executive summary: a report of the American College of Cardiology. / Ryan T.J. [et. al.] // Circulation. 1996. Vol. 94. P. 2341-2350.
2. Arterial gene therapy for therapeutic angiogenesis in patients with peripheral artery disease. / Isner J.M. [et. al.] // Circulation. 1995. Vol. 91. P. 2687-2692.
3. Червяков Ю.В., Власенко О.Н. Эффективность генной терапии и стандартного консервативного лечения хронической ишемии нижних конечностей атеросклеротического генеза. // Вестник хирургии имени ИИ Грекова. 2018. Т. 177. С. 64-71.
4. Kim D.E, Chivian D., Baker D. Protein structure prediction and analysis using the Robetta server. // Nucleic Acids Research. 2004. Vol. 32. P. 526-531.
5. Benkert P., Biasini M., Schwede M. Toward the estimation of the absolute quality of individual protein structure models. // Bioinformatic. 2011. Vol. 27. P. 343-350.

ПРОГНОЗИРОВАНИЕ РЕГУЛЯТОРНЫХ МИШЕНЕЙ ПРОКАРИОТИЧЕСКОЙ РНК С ПОМОЩЬЮ ГЛУБОКОГО МАШИННОГО ОБУЧЕНИЯ

М.А. Сиколенко

Белорусский государственный университет, Минск, Беларусь
E-mail: mishasikolenko@mail.ru

Малые регуляторные РНК широко распространены у прокариотов, причем наиболее изученное семейство из этих не кодирующих генов соответствует транс-действующим регуляторам, которые связываются с их мишенями путем парного взаимодействия оснований. Учитывая увеличивающуюся частоту идентификации этих генов, важно, чтобы методы для выявления их регуляторных целей не отставали. В данной статье исследуются взаимодействия между малыми РНК и их целями с использованием подхода глубокого машинного обучения. Получена метрика качества AUC, равная 0,906. Модель в данной работе обучается на том же наборе данных, что и алгоритм TargetRNA3, и демонстрирует лучшее качество классификации по сравнению с последним.

Ключевые слова: малые регуляторные РНК; нейронные сети; машинное обучение; биоинформатика; глубокое машинное обучение.

АНАЛИЗ НАБОРА ДАННЫХ

Малые РНК широко распространены и играют важную роль в регуляции генов, действуя на посттранскрипционном уровне они связываются с целевыми мРНК, влияя на их трансляцию или стабильность. Они способны воздействовать на несколько мРНК, что позволяет им оказывать широкое влияние на клеточные процессы. Однако их аннотация может быть сложной из-за разнообразия размеров, функций и степени консервативности [2].

В рамках обучающей выборки был использован набор данных для обучения модели TargetRNA3 с целью проведения корректного сравнения эффективности двух моделей. Данный набор данных состоит из 118 колонок. Первые 5 – не численные признаки. В оставшихся 112 колонках содержится пространство признаков, которые были рассмотрены создателями модели TargetRNA3. Кроме того, в наборе данных присутствует целевая переменная «Evinced Interaction», представленная в бинарном формате, указывающая наличие, экспериментальное подтверждение, (единица) или отсутствие (ноль) взаимодействия между малыми РНК и их мишенями. Полный набор данных доступен в файле формата .CSV [1].

ВЫБОР ПРИЗНАКОВ

При работе с обширными наборами признаков необходим качественный отбор, так как модели машинного обучения могут испытывать затруднения в выявлении значимых характеристик. При создании модели TargetRNA3 применялись статистические методы для отбора признаков, в то время как в данной работе применялись методы интеллектуального анализа данных. Изначально в наборе данных используется пространство признаков, включающее результаты работы других алгоритмов, однако для повышения независимости решений от выводов других программ было принято решение сосредоточиться только на характеристиках, специфичных для малых РНК и их мишеней [2].

Для отбора признаков был использован метод «Случайный лес» из библиотеки Scikit-learn. Этот метод основан на построении ансамбля решающих деревьев. Важным для данной работы является поле «feature_importance_» которую содержит обученный случайный лес. Значение данной переменной показывает важность каждого признака в предсказании целевой переменной, позволяя определить, какие из них оказывают наибольшее влияние на модель и являются наиболее информативными.

Результаты анализа показали, что наиболее значимыми признаками являются ACC, GCT и CCT, соответствующие различиям частот тринуклеотидов, а также признаки, связанные с гомологией и длиной последовательности. На основе работы «Случайного леса» было отобрано 45 наиболее важных признаков [3]. Анализ данных показал существенный дисбаланс между классами, что может негативно сказываться на работе моделей машинного обучения. Для уменьшения разницы в классах было решено случайным образом отобрать экземпляры нулевого класса, сократив дисбаланс до разницы в пять раз [3]. Для сокращения размерности пространства признаков был применен метод главных компонент (PCA). Основная цель метода состоит в том, чтобы проецировать исходные признаки на новое пространство меньшей размерности, при этом сохраняя максимально возможное количество дисперсии данных. Пространство признаков было сжато до 35-ти значений [3].

РАЗРАБОТКА АЛГОРИТМА

В рамках задачи машинного обучения решалась задача бинарной классификации. Для решения данной задачи была разработана нейронная сеть с 35 входными нейронами. Количество скрытых слоев и число нейронов на каждом скрытом слое были подобраны с использованием метода

автоматического подбора гиперпараметров. Каждая конфигурация модели была обучена дважды для исключения случайного влияния на хорошее предсказание. После процесса автоподбора была выбрана наилучшая архитектура: 288, 144, 96, 72, 57 и 48 нейронов на последовательно идущих скрытых слоях. В качестве функции активации на скрытых слоях использовалась функция ELU. На выходном слое два нейрона с функцией активации Softmax [3].

СНЯТИЕ МЕТРИК КАЧЕСТВА И СРАВНЕНИЕ С TARGETRNA3

Полученные метрики на тестовом наборе данных представлены в таблице.

Метрики качества Model и TargetRNA3

Алгоритм	AUC	F1	MCC
TargetRNA3	0,75	0,35	0,28
Model	0,90	0,55	0,48

Метрика AUC (Area Under the ROC Curve) представляет собой площадь под кривой ROC (Receiver Operating Characteristic). Чем выше значение AUC, тем лучше качество модели. ROC-AUC кривая строится путем построения графика зависимости True Positive Rate (TPR), вычисляемого по формуле (2), от False Positive Rate (FPR), вычисляемого по формуле (3), при различных порогах классификации.

F1-мера используется как метрика оценки моделей классификации. Она позволяет учитывать, как количество истинно положительных, так и ложноположительных результатов, и вычисляется по формуле (4).

MCC представляет собой коэффициент корреляции между предсказаниями модели и фактическими значениями, учитывая все четыре категории (TP, TN, FP, FN). MCC является более надежной метрикой, чем F1-мера в случае несбалансированных классов, что и наблюдается в данной работе, и вычисляется по формуле (5).

$$TPR = \frac{TP}{TP+FN}, \quad (2)$$

$$FNR = \frac{FP}{TN+FP}, \quad (3)$$

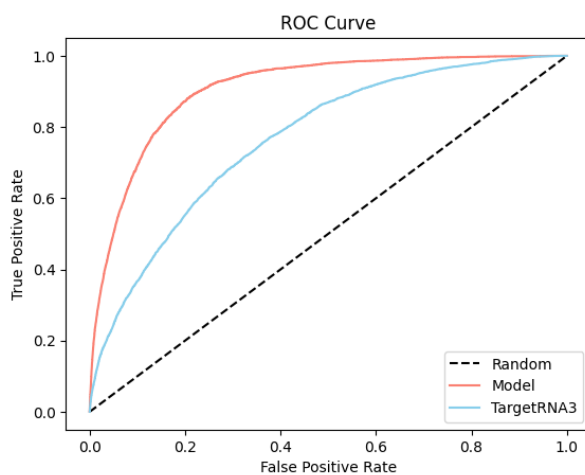
$$F1 = \frac{2*TP}{2TP+FP+FN}, \quad (4)$$

$$MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)*(TP+FN)*(TN+FP)*(TN+FN)}}, \quad (5)$$

где TP – количество истинно положительных предсказаний,

TN – количество истинно отрицательных предсказаний,
 FN – количество ложноотрицательных предсказаний,
 FP – количество ложноположительных предсказаний.

Также данная модель была протестирована на всем наборе данных, и исходя из полученных предсказаний была построена ROC-AUC кривая. Значения TPR и FPR вычисляются согласно формулам (2) и (3). Model – алгоритм, представленный в данной работе. TargetRNA3 – вероятности предсказаний модели TargetRNA3. Метрики AUC на полном наборе данных составляют 0,906 и 0,767 для Model и TargetRNA3 соответственно, рисунок [3].



ROC-AUC кривая

ЗАКЛЮЧЕНИЕ

Несмотря на то, что в работе, посвященной TargetRNA3, одной из моделей-кандидатов была нейронная сеть, ее архитектура и структура остаются неясными. Однако данная работа демонстрирует, что при правильном подборе архитектуры нейронной сети можно достичь значительного улучшения метрик качества и качества классификации регуляторных мишеней прокариотической РНК.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Tjaden B. TargetRNA3: predicting prokaryotic RNA regulatory targets with machine learning. // Harvard Dataverse. 2023. DOI: 10.7910/DVN/2Q8YRF
2. TargetRNA3: predicting prokaryotic RNA regulatory targets with machine learning [Электронный ресурс]. URL: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-023-03117-2> (дата обращения: 17.04.2024).
3. RegulatoryRNAHunter [Электронный ресурс]. URL: <https://github.com/InnerShadow/RegulatoryRNAHunter/tree/main> (дата обращения: 17.04.2024).

БИОИНФОРМАТИЧЕСКИЙ ПАЙПЛАЙН ДЛЯ ОЦЕНКИ ИММУНОГЕННОСТИ БЕЛКОВ ПАТОГЕНОВ И РАКОВЫХ ОПУХОЛЕЙ

И.С. Трусов, Е.В. Гузова, В.В. Гринев

Белорусский государственный университет, Минск, Беларусь
E-mail: ivan.trusau@yandex.by

Данная исследовательская работа посвящена разработке биоинформационного пайплайна для оценки иммуногенности белков. В работе описывается пошаговая сборка пайплайна, а также проводится его сравнение со схожими программными решениями. Разработанный пайплайн превосходит аналоги по точности и F1-score и позволяет в автоматическом режиме генерировать список пептидов целевого белка, ранжированных по иммуногенности.

Ключевые слова: иммуногенность белков; прогнозирование; биоинформатика.

Ежегодно всё больше достижений отмечается в области персонализированной медицины, в частности, благодаря значительному прогрессу в создании персонализированных вакцин, в основе которых лежат генные и дендритно-клеточные технологии [1]. Однако темпы разработки новых вакцин остаются низкими, что является главным препятствием к их массовому и/или персонализированному применению. Одна из причин – проблемы с выявлением иммуногенных пептидов (эпитопов, специфических последовательностей в белках, способных вызывать иммунный ответ). В таком контексте появление биоинформатического инструмента, способного с высокой точностью и чувствительностью предсказывать иммуногенные пептиды, да к тому же ещё и учитывающего различные аспекты процессинга белка, могло бы стать краеугольным камнем при создании вакцин нового поколения.

МАТЕРИАЛЫ И МЕТОДЫ

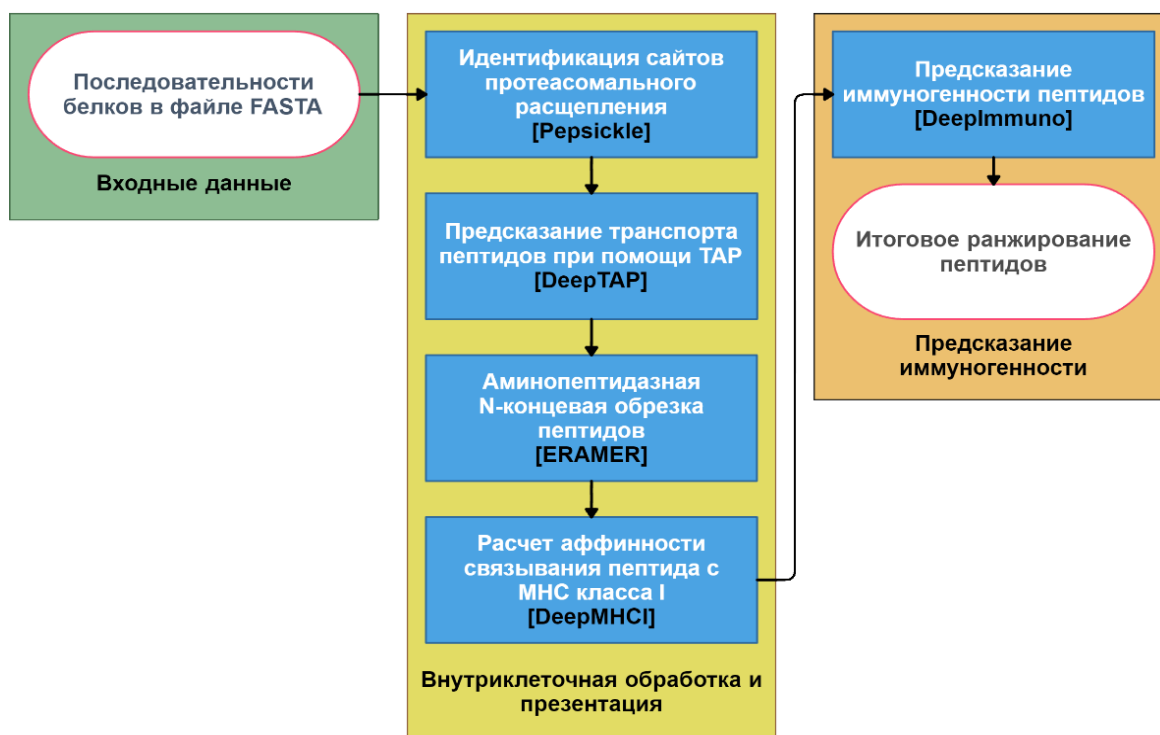
Для проектирования пайплайна были отобраны опубликованные независимые программные пакеты с открытым исходным кодом и наилучшими значениями *AUC* при сравнении с аналогами. Все пакеты были написаны на языке *Python* с применением библиотек *tensorflow*, *pytorch*, *numpy*, *pandas* и *scipy*. Для тестирования пакетов были взяты данные по иммуногенности вирусных пептидов, доступные в базе *IEDB* (англ. immune epitope database) [3]. В тестах учитывались только пептиды, которые имели длину в 9-10 нативных аминокислотных остатков. Иммуногенные свойства таких пептидов были подтверждены экспериментально.

Выбор аллелей главного комплекса гистосовместимости (*MHC*, англ. major histocompatibility complex) соответствовал их использованию в боль-

шинстве исследований. Бенчмаркинг проводился в сравнении с широко известными программными решениями, находящимися в открытом доступе: *NetMHC 4.0*, *NetMHCpan 4.0*, *MHCflurry 2.0* и *MixMHCpred 2.2* [2]. Визуализация результатов осуществлялась с помощью библиотек *Python Matplotlib* и *Seaborn*. Расчёты проводились на компьютере, оснащённом процессором с 8 ядрами и тактовой частотой 3,2 ГГц каждое и 16 Гб оперативной памяти, а также SSD-диском.

СТРУКТУРА И ОРАГНИЗАЦИЯ ПАЙПЛАЙНА

В нашем исследовании мы сосредоточились на внутриклеточном процессинге белков с участием только молекул МНС класса I. Такой процессинг включает в себя пять последовательных этапов: 1) расщепление белков протеасомами, 2) транспортировка полученных пептидов в эндоплазматическую сеть при помощи ТАР (англ. transporter associated with antigen processing), 3) дополнительное расщепление пептидов пептидазами, 4) связывание пептидов молекулами МНС, 5) распознавание комплекса пептида с МНС иммунными клетками. Для анализа этих этапов были выбраны пакеты *pepsickle*, *DeepTAP*, *ERAMER*, *DeepMHC* и *DeepImmuno*, соответственно (см. рисунок).



Графическое изображение структуры и организации пайплайна

Работа пайплайна начинается с *pepsickle*, требующего последовательность белка в формате FASTA. *Pepsickle* возвращает координаты и вероятности разрезания белка протеасомами. Все возможные варианты получаемых

пептидов длиной 9-16 аминокислот передаются DeepTAP, который оценивает вероятность транспорта пептидов в эндоплазматическую сеть. Далее пептиды с длиной более 10 аминокислот дополнительно анализируются пакетом ERAMER, который определяет шансы пептида быть дополнительно расщеплённым ферментом ERAp1. Полученные пептиды длиной 9-10 аминокислот подвергаются анализу с помощью DeepMHCI, который предсказывает вероятность образования комплекса пептида с МНС. Отобранные комплексы затем анализируются с помощью DeepImmuno для ранжирования пептидов по иммуногенности.

В качестве порогового значения вероятностей при отборе пептидов использовались значения 0,5, 0,22, 0,1, 0,44 и 0,88 для pepsickle, DeepTAP, ERAMER, DeepMHCI и DeepImmuno, соответственно. Конечным результатом работы пайплайна является файл формата CSV с отобранными пептидами и краткой информацией (аннотацией) по каждому из этапов обработки.

ОЦЕНКА ПРОИЗВОДИТЕЛЬНОСТИ ПАЙПЛАЙНА

Для оценки эффективности вычислительных алгоритмов в выявлении иммуногенных пептидов было проведено сравнение разработанного нами пайплайна с известными программными решениями. Набор для тестирования включал 63 пептида вирусной природы из 12 исследований, при этом 40 пептидов были не иммуногенны. Каждый из пептидов тестировался в комбинации с 7 разными аллелями МНС. Сравнение проводилось по следующим метрикам: специфичность, чувствительность, метрика *F1* и *MCC* (англ. *Matthews correlation coefficient*).

Оценка пептидов программами *NetMHC*, *NetMHCpan* и *MHCflurry* проводилась путем расчета аффинности связывания пептидов с молекулами МНС по показателю *IC50* (концентрация полумаксимального ингибирования), выраженного в нмоль/л. При этом пептиды считались иммуногенными при стандартном пороговом значении *IC50* в 500 нмоль/л или менее. В программе *MixMHCpred* используется ранговая система оценки пептидов с пороговым значением 0,5. Данное значение было подобрано эмпирически – при нём демонстрировалась наилучшая точность предсказаний. Для нашего пайплайна пептиды считались иммуногенными, если они прошли все этапы и были представлены в конечном файле CSV.

Сравнительный анализ показал, что разработанный нами пайплайн демонстрирует лучшие результаты по сравнению с аналогами по показателям специфичности, метрике *F1* и *MCC* (см. таблицу). Очевидно, что наш пайплайн уступает аналогам только по такому показателю, как чувствительность: отношению количества обнаруженных программой событий к их истинному количеству. Однако этот показатель связан со специ-

фичностью, и снижение специфичности может повышать чувствительность. А поскольку для разработки эффективных вакцин более важным является показатель специфичности (для успешной вакцины достаточно, как правило, нескольких высокоиммуногенных пептидов), то мы не считаем более низкую чувствительность нашего пайплайна его существенным недостатком.

Статистические результаты тестирования программ

Метрика	<i>NetMHC</i>	<i>NetMHCpan</i>	<i>MHCflurry</i>	<i>MixMHCpred</i>	пай-плайн
Специфичность	0,43	0,47	0,39	0,52	0,84
Чувствительность	0,83	0,91	1,00	0,70	0,70
<i>F1</i>	0,57	0,62	0,56	0,59	0,76
<i>MCC</i>	0,21	0,33	0,20	0,31	0,65

ЗАКЛЮЧЕНИЕ

Результатом нашего исследования стал биоинформатический пайплайн, позволяющий идентифицировать иммуногенные пептиды в изучаемом белке. Он основан на предсказательных моделях и программных пакетах, объединённых с помощью *Python*. Модульная архитектура пайплайна облегчает его настройку и интеграцию дополнительных функций, что делает пайплайн хорошей базой для дальнейших исследований. В тестах по оценке иммуногенности вирусных белков наш пайплайн продемонстрировал высокую специфичность, а также оказался лучше своих аналогов по метрикам *F1* и *MCC*. Полученные результаты указывают на перспективность дальнейшего развития данного подхода.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. IEDB and CEDAR: two sibling databases to serve the global scientific community. / Blazeska N., Kosaloglu Z.Y., Vita R. [et. al.] // Computational Vaccine Design. 2023. Vol. 2673. P. 133-149. DOI: 10.1007/978-1-0716-3239-0_9
2. Трусов И.С. Биоинформатическое прогнозирование иммуногенности генно-инженерных белков для активной иммунотерапии // Матер. 80 науч. конф. студ. и аспирантов. БГУ, 2023. С. 132-136.
3. Personalized dendritic cell vaccines – recent breakthroughs and encouraging clinical results. / Mastelic B.G., Balint K., Boudousquie C. [et. al.] // Cancer Immunity and Immunotherapy. 2019. Vol. 10. P. 1-6. DOI: 10.3389/fimmu.2019.00766

ИДЕНТИФИКАЦИЯ СОВПАДАЮЩИХ ФРАГМЕНТОВ НЕКОДИРУЮЩИХ УЧАСТКОВ МИТОХОНДРИАЛЬНЫХ ГЕНОМОВ НАСЕКОМЫХ ОТРЯДА ПОЛУЖЕСТКОКРЫЛЫЕ (ЛАТ. HEMIPTERA)

**В.И. Чесалин, Н.В. Воронова-Барте, С.С. Левыкина, В.Л. Крук,
Н.С. Мешич, П.Е. Александрович**

*Белорусский государственный университет, г. Минск, Беларусь
E-mail: vitaliy.krukh@gmail.com, nikitameshich@gmail.com*

Разработан алгоритм построения множества максимальных по включению общих фрагментов в некодирующих участках митохондриальных геномов насекомых.

Ключевые слова: митохондриальные геномы; молекулярная генетика; насекомые.

Введение. Митохондриальные геномы насекомых представляют собой кольцевую молекулу ДНК, размером от 15 до 18 kb. Они содержат 13 белок-кодирующих генов, 22 гена митохондриальных тРНК и 2 гена рРНК (12S-рРНК и 16S-рРНК). Как и у других животных, в митохондриальных геномах насекомых отсутствуют интроны и протяженные межгенные спейсеры. Однако в структуру мтДНК входят две некодирующие области: Repeat-регион, или область тандемных повторов, встречающийся только у части видов; а также обязательный участок – Control-region, или область формирования D-петли [1].

Repeat-регион – участок митохондриальных геномов, длиной в около 1000 п.н., который содержит длинные, тандемно расположенные повторы [2]. Наличие такого крупного участка в митохондриальных геномах представляет большой интерес, поскольку эволюция мтДНК идет по пути компактизации. В настоящее время функциональное значение региона-повторов, а также его происхождение не известно [3]. Control-регион, или область формирования D-петли – участок митохондриальных геномов, представляющий собой АТ-богатую область, функциональное назначение которого – инициация репликации ДНК [4,5]. Известно, что положение в митохондриальном геноме и длина области формирования D-петли переменны, что обусловлено наличием в них тандемных повторов, функциональное значение которых остается дискуссионным [4].

В нашей работе мы сделали попытку установить наличие или отсутствие структурного сходства между repeat- и control-регионами, приняв в качестве гипотезы утверждение, что обнаруженное сходство в мотиве нуклеотидных последовательностей будет указывать на общность происхождения repeat-региона и D-петли. Для проверки нашей гипотезы мы по-

ставили перед собой задачу разработать алгоритм для обнаружения областей подобия с возможностью проведения множественного попарного сравнения между видами и между таксонами (в прямой, обратной и комплементарной ориентациях), а также с возможностью устанавливать нужный процент совпадений, поскольку на данный момент не существует оптимального пакета программ, позволяющего быстро и эффективно выполнить наш анализ с минимальными затратами ресурсов.

Материалы и методы. Для проверки гипотезы была создана выборка из 262 нуклеотидных последовательностей полных мтДНК насекомых, депонированных в открытом доступе в базе данных GenBank. Из полных геномов были извлечены последовательности, соответствующие областям control- и repeat-регионов. Сводные данные о составе выборки представлены в таблице.

Таксономическая выборка

Таксон насекомых	Количество геномов		Суммарное количество геномов
	С repeat-регионом	Без repeat-региона	
Coccoidea	3	4	7
Aleyrodidae	4	6	10
Psylloidea	4	32	36
Cicadellidae	1	57	57
Heteroptera	10	84	94
Aphididae	35	18	53

Для каждого таксона T с контрольной областью C и региона повторов R рассмотрим множество общих фрагментов последовательностей

$$X_T = X_C \cap X_R, \quad (1)$$

где X_C – множество всех фрагментов последовательностей из контрольной области C и X_R – множество всех фрагментов последовательностей из региона повторов R . На множестве X_T введем отношение порядка $\preceq$ следующим образом: пусть

$$a, b \in X_T, b = (b_1, b_2, \dots, b_n) \quad (2)$$

Тогда полагаем

$$a \preceq b \leftrightarrow \exists(i, j), 1 \leq i \leq j \leq n, a = (b_i, b_{i+1}, \dots, b_j) \quad (3)$$

Таким образом, пара $(X_T, \preceq)$ образует частично упорядоченное множество, максимальные элементы которого характеризуют степень сходства между последовательностями C и R . Для анализа сходства последовательностей мы используем алгоритм, изображенный на рис. 1. Данный алгоритм можно разделить на 2 этапа:

1. Создание множеств фрагментов для каждой последовательности. Поиск их пересечения.
2. Поиск множества максимальных элементов.
 - 2.1. Создание сортированного по длине последовательностей списка S на основе пересечения множеств фрагментов.
 - 2.2. Создание пустого списка M_S для хранения максимальных элементов.
 - 2.3. Итеративный проход по отсортированному множеству строк. Для каждой строки из множества S проверяем, входит ли она в некоторую строку из M_S . Если нет, добавляем её в M_S .

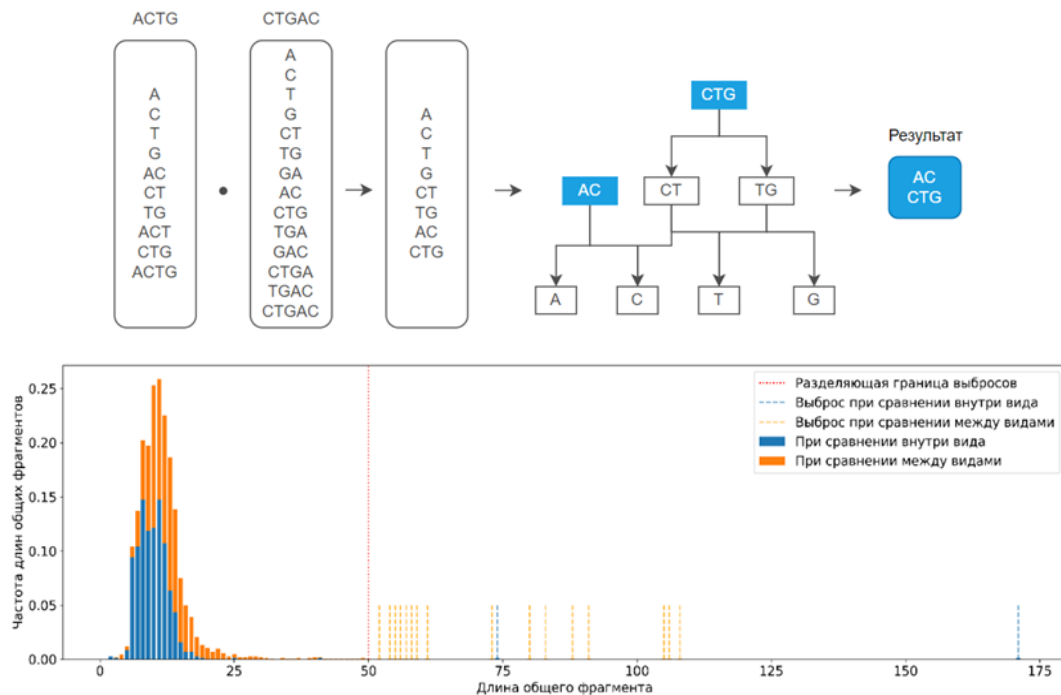


Рис. 1. Иллюстрация работы алгоритма

Результаты. Несмотря на установление жесткого порога для сходства в 100 % и минимальной длины последовательностей в 50 нуклеотидов (полагая, что при 100 % совпадении последовательностей длиной 50 нуклеотидов и выше мы можем исключить вероятность случайного совпадения) были обнаружены 2 случая, в которых в control- и gеreat-регионах присутствовали идентичные участки длиной 279 нуклеотидов в геноме Aphididae (рис. 2) и три участка длиной по 185, 199 и 239 нуклеотидов в одном геноме Heteroptera (рис. 3). Наличие столь протяженных идентичных участков в двух отдельных областях генома, являющихся некодирующими, едва ли является случайным. Это не позволяет нам утверждать, что gереат-регион был образован в результате дупликации и транслокации участков control-региона, однако дальнейшие исследования с установлением меньшего порога сходства, вероятно, позволят нам ответить на этот вопрос.

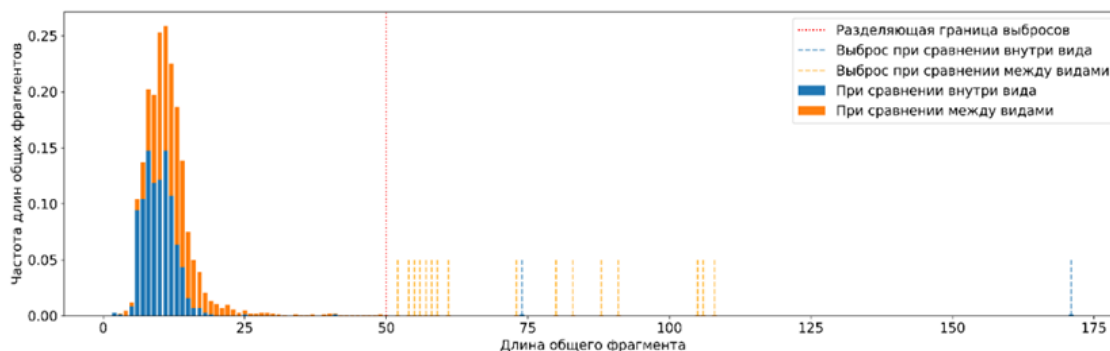


Рис. 2. Сгруппированная столбчатая гистограмма длин максимальных фрагментов таксона Aphididae.

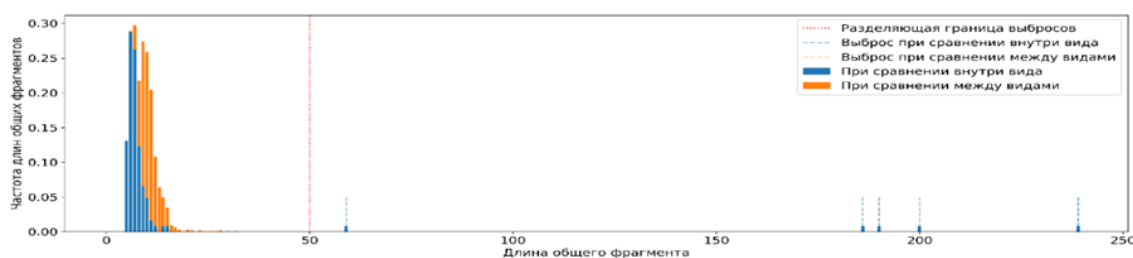


Рис. 3. Сгруппированная столбчатая гистограмма длин максимальных фрагментов таксона Heteroptera.

Разделяющая граница выбросов – величина, задающая формальный критерий того, что мы считаем выбросом: выбросами считаются все величины, которые больше данной границы.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. The mitochondrial genome of *Greenidea psidii* van der Goot (Hemiptera: Aphididae: Greenideinae) and comparisons with other Aphididae aphids. / Chen J. [et al.] // International Journal of Biological Macromolecules. 2019. Vol. 122. P. 824-832.
2. Воронова Н.В., Левыкина С.С., Бондаренко Ю.В., Шулинский Р.С. Некодирующие области в митохондриальном геноме тли *Aphis fabae mordvilkoï* Börner & Janisch, 1922. // Молекулярная и прикладная генетика. 2019. Т. 26. С. 64-72.
3. Characteristic and variability of five complete aphid mitochondrial genomes. / Voronova N.V. [et al.] // International Journal of Biological Macromolecules. 2020. Vol. 149. P. 187-206.
4. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. / Zhang D.X. [et al.] // Biochemical Systematics and Ecology. 1997. Vol. 25. P. 99-120.
5. A novel closed-circular mitochondrial DNA with properties of a replicating intermediate. / Kasamatsu H. [et al.] // PNAS. 1971. Vol. 68. P. 2252-2257.

ФАРМАКОФОРНАЯ МОДЕЛЬ НА ОСНОВЕ СТРУКТУРЫ ЛИГАНДА CYP17A1 ЧЕЛОВЕКА

М.И. Шаладонова, Я.В. Диченко, С.А. Усанов

Институт биоорганической химии НАН Беларуси, Минск, Беларусь
E-mail: shalmari@tut.by

Работа посвящена созданию фармакофорной, 3D-QSAR моделей на основе структуры лигандов CYP17A1 человека, селективно ингибирующих 17,20-лиазную активность. С применением библиотек Python 3.11 определены основные фармакофорные признаки, обуславливающие биологическую активность химических соединений. При помощи программы Schrodinger построена фармакофорная модель и 3D-QSAR модель для определения IC_{50} по отношению к 17,20-лиазной активности CYP17A1.

Ключевые слова: лиганды 17,20-лиазы; фармакофор; фармакофорные признаки; 3D-QSAR.

Фермент CYP17A1 (17 α -гидроксилаза /17,20-лиаза) относится к группе цитохромов суперсемейства P450 и участвует в реакциях биосинтеза глюкокортикоидов и андрогенов. В то же время CYP17A1 человека вовлечен в патогенез такого заболевания, как рак предстательной железы (РПЖ). Единственным в мире зарегистрированным препаратом-ингибитором CYP17A1 для лечения РПЖ является абиратерона ацетат, который представляет собой стероидный ингибитор, имитирующий природный субстрат фермента. Стоит отметить, что преимущественное ингибирование 17 α -гидроксилазной активности фермента, свойственное абиратерону, приводит к дефициту глюкокортикоидов и синдрому избытка минералкортикоидов, что сопровождается значительными побочными эффектами. Чтобы исключить данные побочные эффекты, необходимо применение ингибиторов, которые будут влиять преимущественно на 17,20-лиазную активность фермента. С данной целью ведется поиск новых селективных к 17,20-лиазе стероидных и нестероидных ингибиторов CYP17A1. В настоящее время для оптимизации поиска потенциальных лигандов используют методы компьютерного моделирования, позволяющие сократить время и затраты на высокопроизводительный лабораторный скрининг. Фармакофорное моделирование является эффективным методом отбора биологически активных химических соединений на первом этапе компьютерного поиска лекарственных веществ.

Цель исследования: при помощи компьютерных методов фармакофорного анализа получить фармакофорную модель для ингибиторов 17,20-лиазной активности CYP17A1 и предложить возможность использования фармакофорной модели для скрининга новых лигандов, перспективных для изучения *in vitro* и *in vivo*. Для достижения поставленной цели необходимо решить следующие задачи: проанализировать научные статьи

с экспериментальными результатами *in vitro* для выявления наиболее эффективных ингибиторов 17,20-лиазной активности, найти общие фармакофорные признаки для ингибиторов, провести фармакофорное и 3D-QSAR моделирование.

ОСНОВНАЯ ЧАСТЬ

Фармакофорное и 3D-QSAR моделирование проводили с использованием инструментария программы Schrodinger (модули Phase, LigPrep, MacroModel, Epik). Для фармакофорного моделирования использовали лиганды, изученные на рекомбинантном CYP17A1 человека *in vitro*, для которых значения IC₅₀ относительно 17,20-лиазной активности составляют от 2 до 38 нмоль. Базу данных лигандов CYP17A1 человека формировали на основании информации, представленной в научных статьях. Количество соответствующих соединений составило 23 лиганда нестероидной структуры, представляющих наиболее эффективные ингибиторы. Для выбранных соединений были получены 2D структуры из базы данных ChemSpider (<https://www.chemspider.com>), которые конвертировали в 3D структуры при помощи модуля LigPrep с параметрами силового поля OPLS-4 при pH 7,4 с использованием модуля Epik, учитывая таутомерные формы и стереоизомеры. Затем 3D структуры подвергали конформационному поиску с использованием алгоритма модуля MacroModel (силовое поле OPLS-4, метод систематического поиска с последующей минимизацией энергии полученных конформаций, константа диэлектрической проницаемости 1,0, показатель среднеквадратичного отклонения не более 1 Å). Среднее количество конформеров для каждой молекулы составило 39 (от 26 до 50 конформаций для одного соединения). С учетом стереоизомеров, таутомеров и конформеров, стабильных при pH 7,4 получено 900 структур соединений.

Для поиска фармакофорных признаков полученных соединений применяли библиотеки Python 3.11 v.2024.2.1. (pandas v.2.2.1, rdkit v. 2023.9.5, matplotlib v. 3.8.2). Количество фармакофорных признаков для активных химических соединений представлено в табл. 1.

Таблица 1

Количество фармакофорных признаков для активных соединений

Фармакофорные признаки	Количество
Акцепторы водородной связи	От 2 до 4
Доноры водородной связи	От 3 до 5
Катионные центры	1
Анионные центры	Отсутствуют
Гидрофобные группы	От 3 до 15
Ароматические циклы	3

При использовании модуля Phase программы Schrodinger проводили компьютерное моделирование для получения общей фармакофорной модели, причем применяли параметр полного соответствия фармакофору как минимум для 90% из полученных 900 соединений. Построенная общая фармакофорная модель с наилучшими скоринговыми показателями представлена на рис. 1 и включает в себя два гидрофобных участка (зеленые сферы), два ароматических цикла (оранжевые кольца) и один донор водородной связи (синяя сфера). Найденные фармакофорные признаки лигандов, используемых для построения модели, принимают непосредственное участие в связывании с активным центром фермента и обуславливают ингибирование 17,20-лиазной активности CYP17A1.

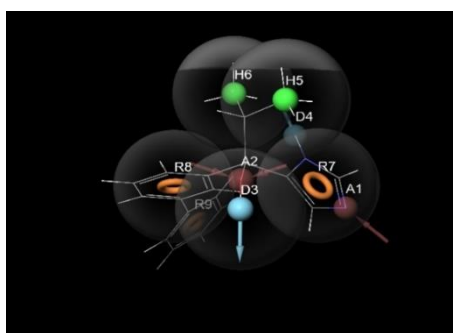


Рис. 1. Общая фармакофорная модель на основе структуры лигандов CYP17A1

С применением функционала модуля Phase строили 3D-QSAR модели при помощи метода регрессии частичных наименьших квадратов (PLS). Алгоритмические модели обучали предсказывать IC_{50} для соединений. Исходную выборку лигандов делили стохастическим методом на обучающую (80%) и тестовую (20%). Статистические параметры, позволяющие определить качество полученных моделей, представлены в табл. 2. Исходя из вычисленных параметров, наилучшей является модель 5.

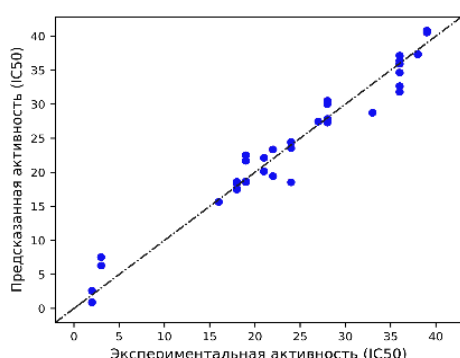
Графики линейной регрессии 3D-QSAR модели для химических соединений обучающей и тестовой выборки представлены на рис. 2.

ЗАКЛЮЧЕНИЕ

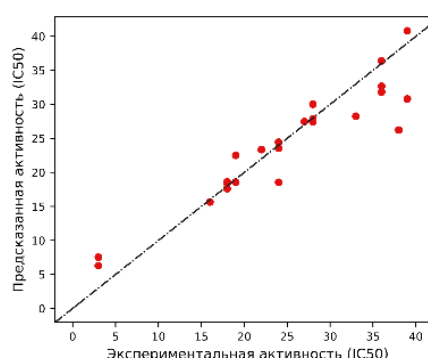
В результате фармакофорного моделирования с применением компьютерной программы Schrodinger и языка программирования Python 3.11 построена фармакофорная и 3D-QSAR модели для поиска новых ингибиторов CYP17A1, селективно ингибирующих 17,20-лиазную активность. Полученная модель может быть пригодной для использования на этапах скрининга обширных баз данных химических веществ с целью выявления новых нестероидных ингибиторов 17,20-лиазы человека.

Статистические показатели 3D-QSAR моделей

Модель	Коэффициент детерминации для обучающей выборки (R^2)	Коэффициент детерминации для тестовой выборки (Q^2)	Коэффициент корреляции Пирсона (Pearson-r)
Модель 1	0,7390	0,5029	0,7191
Модель 2	0,8448	0,6663	0,8429
Модель 3	0,9016	0,7519	0,8898
Модель 4	0,9480	0,8095	0,9292
Модель 5	0,9674	0,8298	0,9316



а



б

Рис 2. График линейной регрессии для экспериментальных и предсказанных значений IC_{50} для обучающей (а) и тестовой (б) выборок

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Диченко Я.В. Компьютерное моделирование строения и реакционной способности молекул. // Минск: Белорусская наука, 2023. С. 139.
2. Prospective computational design and in vitro bio-analytical tests of new chemical entities as potential selective CYP17A1 lyase inhibitors. / Gumede N.J., Nxumado W., Bissetty K. [et al.] // Bioorganic chemistry. 2019. Vol. 94, № 103462. P. 1-16. DOI: 10.1016/j.bioorg.2019.103462
3. Non-steroidal CYP17A1 inhibitors: discovery and assessment. / Wróbel T.M., Jørgensen F.S., Pandey A.V. [et al.] // Journal of medicinal chemistry. 2023. Vol. 66, № 10. P. 6542–6566. DOI: 10.1021/acs.jmedchem.3c00442
4. C-(17,20)-lyase inhibitors. Part 2: design, synthesis and structure-activity relationships of (2-naphthylmethyl)-1H-imidazoles as novel C-(17,20)-lyase inhibitors. / Matsunaga N., Kaku T., Ojida A. [et al.] // Bioorg Med Chem. 2004. Vol. 12, № 16. P. 4313-4336. DOI: 10.1016/j.bmc.2004.06.016
5. 17,20-Lyase inhibitors. Part 4: Design, synthesis and structure-activity relationships of naphthylmethylimidazole derivatives as novel 17,20-lyase inhibitors. / Kaku T., Matsunaga N., Ojida A. [et al.] // Bioorganic & Medicinal Chemistry. 2011. Vol. 19, № 5. P. 1751-1770. DOI: 10.1016/j.bmc.2011.01.017

ГЕНЫ СИСТЕМЫ ДЕТОКСИКАЦИИ В ГЕНОМЕ ТЛИ *MACROSIPHUM ROSAE*

Н.В. Воронова-Барте, С.С. Левыкина, А.М. Шульгович

Белорусский государственный университет, Минск, Беларусь

E-mail: anyashulgovich@gmail.com

В работе представлены результаты аннотации ABC-транспортеров системы детоксикации в геноме *Macrosiphum rosae*. В геноме *M. rosae* найдено 67 генов ABC-транспортеров. Средняя длина гена – 5755 нуклеотидов, средняя длина экзонов – 1846 нуклеотидов, средняя длина интронов – 3909 нуклеотидов.

Ключевые слова: геномика; аннотация генома; тли.

ВВЕДЕНИЕ

Система детоксикации у насекомых включает в себя монооксигеназы цитохрома P450 (CYP450), глутатион S-трансферазы (GST), карбоксил-эстеразы (CSE), гликозилтрансферазы (UDT) и АТФ-связывающие каскадные транспортеры (ABC). Белки системы детоксикации метаболизируют не только защитные вещества растений, но также обеспечивают адаптацию насекомого к инсектицидам. Возникновение устойчивости у таких фитофагов может быть обусловлено увеличением копий этих генов и тем, что копии становятся уникальными для каждого метаболита.

ABC-транспортеры это суперсемейство интегральных мембранных белков, которые используют АТФ для перемещения различных субстратов через липидные мембраны. Они играют основную роль в заключительной фазе процесса детоксикации ксенобиотиков. В связи с этим, идентификация их генов является важной для разработки эффективных средств контроля численности, например, таких сельскохозяйственных вредителей как тли.

МАТЕРИАЛЫ И МЕТОДЫ

Целью исследования являлась аннотация генов ABC-транспортеров системы детоксикации *Macrosiphum rosae* фауны Беларуси.

Предварительно была проведена функциональная аннотация генома *Macrosiphum rosae*. Финальная модель была получена как консенсус между НММ-предсказаниями (скрытые марковские модели), белковой и РНК-гомологией, что подтверждает функциональность обнаруженных генов. Получение консенсуса проводилось в программе Evidence Modeller [1].

Поиск доменов в полученных моделях генов проводилась с помощью поиска IPR и Pfam-доменов в Interproscan и интерпретируемого языка про-

граммирования Python [2]. Каждый обнаруженный ген *M. rosae* был про-
 курирован в геномном браузере Apollo для уточнения границ экзонов: рас-
 положения старт- и стоп-кодонов в последовательностях, канонических
 сайтов сплайсинга [3].

РЕЗУЛЬТАТЫ

Результаты структурно-функционального описания идентифициро-
 ванных генов представлены в таблице.

Структурно-функционального описания идентифицированных генов

Семейство	Идентификационный номер	Локус (ориентация)	Длина гена, п.н.	Число экзонов	Длина экзонов, п.н.
A5	MACRO000011987	Scaffold_168444 (-)	3355	6	1650
B6	MACRO000000126	Scaffold_163101 (+)	7166	13	2523
C	MACRO000001134	Scaffold_168689 (-)	10796	20	3612
F2	MACRO000003103	Scaffold_165388 (+)	5353	11	1725
F4	MACRO000015501	Scaffold_152023 (-)	1198	2	1014
G4	MACRO000014479	Scaffold_156322 (-)	1825	6	663
G4	MACRO000005456	Scaffold_166244 (+)	8210	8	1356
G4	MACRO000016283	Scaffold_183958 (-)	1234	2	435
G4	MACRO000002655	Scaffold_185181 (+)	9098	12	1984
G8	MACRO000004229	Scaffold_169484 (+)	2339	9	1563
G9	MACRO000003028	Scaffold_153724 (+)	5314	13	1827
G20	MACRO000017447	Scaffold_159883 (+)	2745	2	447
G20	MACRO000002548	Scaffold_167411 (-)	3717	12	1709
G20	MACRO000002423	Scaffold_171574 (+)	12921	12	1512
G23	MACRO000000434	Scaffold_171604 (+)	9280	14	1833
G23	MACRO000000433	Scaffold_171604 (+)	5192	11	1479
G23	MACRO000003467	Scaffold_152778 (-)	2537	3	492
G23	MACRO000000996	Scaffold_153633 (-)	6610	15	2262
G23	MACRO000011734	Scaffold_157450 (+)	2246	3	414
G23	MACRO000005003	Scaffold_153904 (-)	2109	8	1067
G23	MACRO000008455	Scaffold_164900 (+)	1490	6	882
G23	MACRO000003449	Scaffold_168768 (-)	6732	10	1308
G23	MACRO000015911	Scaffold_207302 (+)	368	2	291
G23	MACRO000001104	Scaffold_165944 (+)	3774	11	1728

Продолжение табл.

G23	MACRO00002385	Scaffold_151776 (-)	7726	12	1530
G23	MACRO00002271	Scaffold_166891 (+)	7448	14	1947
H2	MACRO000015598	Scaffold_184754 (-)	259	2	198
mdr	MACRO00005670	Scaffold_157806 (+)	7493	22	3801
mdr 1	MACRO00001623	Scaffold_156325 (-)	7517	15	2409
mdr1	MACRO00005499	Scaffold_161591 (-)	3314	7	1173
mdr1	MACRO00006419	Scaffold_161800 (-)	2112	6	906
mdr1	MACRO00005876	Scaffold_164422 (+)	2621	8	1197
mdr1	MACRO00000343	Scaffold_164804 (-)	22212	28	4503
mdr 1	MACRO00002636	Scaffold_168537 (-)	6074	12	2061
mdr1	MACRO00005006	Scaffold_169760 (+)	4228	8	1212
mdr1	MACRO00003710	Scaffold_152153 (+)	2514	10	1812
mdr1	MACRO000017853	Scaffold_152132 (-)	2351	2	621
mdr 1A	MACRO00000777	Scaffold_170303 (+)	6600	21	3279
mdr1A	MACRO00002215	Scaffold_166270 (+)	3986	13	1995
mdr4	MACRO00000682	Scaffold_162556 (-)	19295	22	3840
mdr4	MACRO00000601	Scaffold_166690 (-)	7009	24	4026
mdr4	MACRO00000157	Scaffold_159141 (-)	7748	22	3798
mdr4	MACRO00000156	Scaffold_159141 (-)	11795	13	1794
mdr4	MACRO00003994	Scaffold_159646 (+)	4207	10	1548
mdr4	MACRO00000696	Scaffold_164506 (-)	5604	22	3714
mdr4	MACRO00000616	Scaffold_167669 (+)	8385	23	4017
mdr4	MACRO00000117	Scaffold_168154 (+)	23246	38	6378
mdr7	MACRO000014665	Scaffold_165398 (+)	852	3	711
mdr7	MACRO00007471	Scaffold_167420 (-)	1520	6	1080
mdr lethal	MACRO00005131	Scaffold_170856 (-)	1932	8	1128
mdr lethal	MACRO00000845	Scaffold_168251 (-)	10125	20	3420
mdr lethal	MACRO00000630	Scaffold_168346 (+)	12493	23	3951
sub A1	MACRO000011734	Scaffold_157494 (-)	12781	14	4080
subA3	MACRO00007998	Scaffold_163778 (-)	3557	5	1122
subA3	MACRO00004182	Scaffold_167170 (+)	9783	10	3177
sub A3	MACRO000011734	Scaffold_157611 (-)	4968	9	1866
subA5	MACRO00004861	Scaffold_161208 (+)	3700	9	2229

Окончание табл.

subB7	MACRO00005514	Scaffold_171561 (+)	7830	7	1236
subB8	MACRO00003066	Scaffold_165543 (+)	4350	11	1620
sub B10	MACRO00005670	Scaffold_157806 (+)	3343	7	1428
subD3	MACRO000011458	Scaffold_171034 (+)	1943	4	990
subE1	MACRO00008390	Scaffold_165801 (-)	2244	6	1803
subF1	MACRO00004295	Scaffold_169697 (-)	3374	9	1878
subG1	MACRO00004639	Scaffold_167900 (+)	8997	8	1461
subG1	MACRO00004814	Scaffold_168657 (-)	2430	8	1161
subG4	MACRO00003560	Scaffold_161549 (-)	3843	10	1566
subG4	MACRO00005106	Scaffold_169912 (-)	2796	9	1242

В результате было найдено 67 генов ABC-транспортеров. 25 из них относятся к семейству G, 3 – к семейству F, 1 – к семейству H, 1 – к семейству C, 1 – к подсемейству D, 6 – к подсемейству A, 1 – к подсемейству E, 8 – к генам множественной устойчивости 4 (multidrug resistance-associated protein 4), 4 – к летальным генам множественной устойчивости (probable multidrug resistance-associated protein lethal), 11 – к генам множественной устойчивости 1 (multidrug resistance-associated protein 1), 2 – к генам множественной устойчивости 7-подобные (multidrug resistance-associated protein 7-like), 4 – к подсемейству B.

Средняя длина гена – 5854 нуклеотидов, средняя длина экзонов – 1921 нуклеотидов, средняя длина интронов – 3933 нуклеотидов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the program to assemble spliced alignments. / Haas B.J., Salzberg S.L., Zhu W. [et al.] // Genome Biology. 2008. Vol. 9. P. R7. DOI: 10.1186/gb-2008-9-1-r7
2. InterProScan 5: genome-scale protein function classification. / Jones P., Binns D., Chang H., [et al.] // Bioinformatics. 2014. Vol. 30, № 9. P. 1236-1240.
3. Web Apollo: a web-based genomic annotation editing platform. / Lee E., Helt G., Reese J. [et al.] // Genome Biology. 2013. Vol. 14, № 8. P. R93. DOI: 10.1186/gb-2013-14-8-r93

СЕКЦИЯ «ПРИКЛАДНОЙ АНАЛИЗ ДАННЫХ»

ИССЛЕДОВАНИЯ ЭЛЕКТРОМАГНИТНОГО КАНАЛА УТЕЧКИ ИНФОРМАЦИИ СРЕДСТВ ВЫЧИСЛИТЕЛЬНОЙ ТЕХНИКИ

Xiong Shuai, В.Э. Яскевич

Белорусский государственный университет, г. Минск, Беларусь

E-mail: Yaskevich_VE@bsu.by

Исследования посвящены оценке угрозы утечки информации, обрабатываемой современными вычислительными средствами, по электромагнитному каналу. Проблема известна десятки лет. Однако, раньше так называемые TEMPEST атаки были уделом дорогостоящих средств иностранной технической разведки. Развитие современных приемных устройств, программно-определяемых приемников (SDR), анализаторов спектра реального времени сделало такую атаку гораздо доступнее. Это может существенно изменить модель нарушителя и перечень угроз как для информационных систем, так и для объектов информатизации. В докладе представлены методика исследований электромагнитного канала утечки информации средств вычислительной техники и результаты экспериментальных исследований.

Ключевые слова: электромагнитный канал утечки информации средств вычислительной техники.

ВВЕДЕНИЕ

Формирование электромагнитного канала утечки информации можно считать завершенным при обеспечении условий размещения приемника, его технических характеристик и параметров, а также выборе наиболее подходящей антенной системы. В последнее время появилось ряд публикаций, описывающих формирование электромагнитного канала утечки информации для VGA интерфейса на основе общедоступных устройств, включая программно-определяемые радиоприемники (SDR) [1, 2, 3]. Однако все они представляют попытку прямого представления изображения монитора, не учитывая потенциальные возможности последующей обработки зарегистрированных сигналов. Представленные исследования направлены на оценку степени угрозы утечки информации по электромагнитному каналу при использовании анализаторов спектра реального времени, выполненных в виде USB приставок, позволяющих записывать IQ сигналы электромагнитных излучений мониторов.

МЕТОДИКА ИССЛЕДОВАНИЙ

В качестве источника побочных электромагнитных излучений (ПЭМИ) VGA интерфейса использовался настольный компьютер с ЖК-

монитором. Излучателем электромагнитных волн в этом случае, в основном, являлся VGA кабель.

Магнитная антенна для приема ПЭМИ была изготовлена из медной трубки, одного витка провода и согласующего трансформатора 9:1. Выбор магнитной антенны обусловлен сложной электромагнитной обстановкой в месте проведения исследований: условия города и учебного корпуса с множеством лабораторий.

В качестве приемных устройств использовались анализатор спектра реального времени BB60C (The Signal Hound) и RSA306B (Tektronix). Оба анализатора являются USB приставками и требуют высокоскоростного интерфейса и управляющих программ.

Анализатор спектра BB60C работал под управлением программы RadioInspectorRT (ООО «РадиоСофт», г. Москва), реализующей функции радиомониторинга и имеющей в своем составе детектор аналогового телевизионного сигнала. Именно в качестве аналогового телевизионного приемника с возможностью ручной синхронизации он и использовался.

Анализатор спектра RSA306B работал под управлением штатной программы Tektronix SignalVu-PC. Этот анализатор использовался для регистрации IQ сигналов ПЭМИ VGA интерфейса с целью последующей обработки полученных записей.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

На рис. 1 представлены результаты регистрации ПЭМИ VGA кабеля с помощью анализатора спектра BB60C и программного обеспечения RadioInspectorRT, работающего в режиме детектора аналоговых телевизионных сигналов. На экране исследуемого монитора специальным программным обеспечением формировалось шесть чередующихся белых и черных вертикальных полос, одинаковой ширины, равномерно заполняющих экран (рис. 1 а) представляет принятый сигнал от черного экрана. Даже при отсутствии изображения работают сигналы синхроимпульсов и приемное устройство удалось синхронизировать с помощью ручного управления. На рис. 1 б) мы видим несколько отображений исследуемого экрана с вертикальными полосами. Отображение нескольких экранов, а также наклон вертикальных полос обусловлен различной частотой строчной и кадровой разверток исследуемого монитора и стандарта NTSC, который реализуется TV детектором приемного устройства. Ниже изображения экрана на рис. 1 представлен спектр радиосигнала при черном экране и вертикальных полосах. Красная линия (так называемый порог обнаружения) занимает на обоих рисунках а) и б) неизменное положение и служит для визуальной оценки увеличения уровня сигнала при отображении полос на экране.

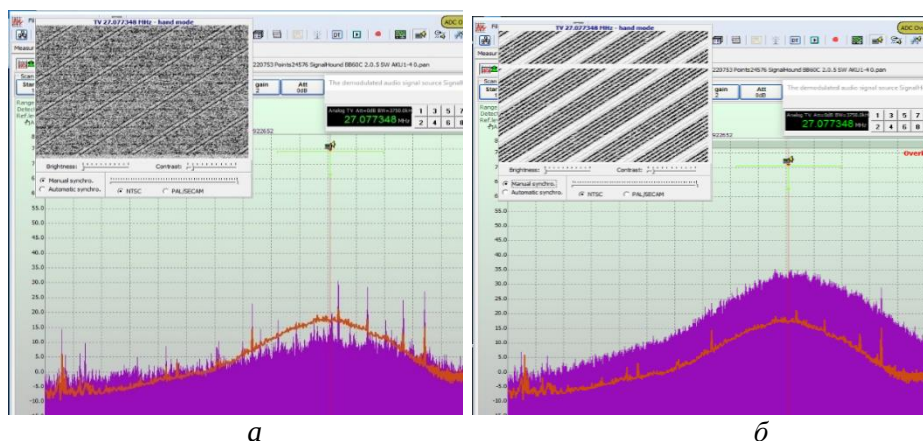


Рис. 1. Регистрация изображения монитора детектором телевизионных сигналов:
а - черный экран; б - вертикальных полос

На рис. 2 представлен прием ПЭМИ VGA интерфейса при отображении на исследуемом мониторе слова «PEACE» с размером шрифта 300 пикселей. Различие в кадровых и строчных частотах привело к разрыву слова и смещения конца слова в начало.

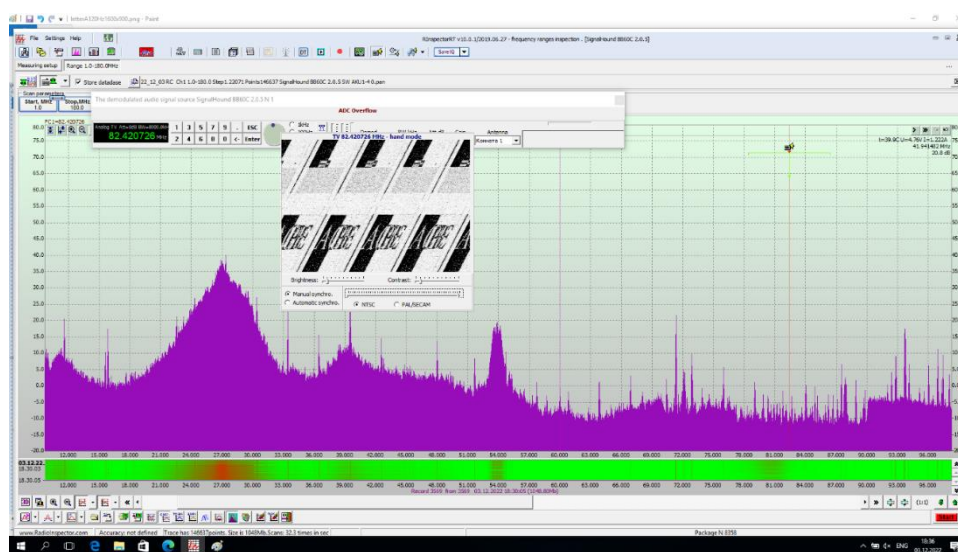


Рис. 2. Регистрация изображения слова PEACE

На рис. 3 представлен фрагмент записи сигналов ПЭМИ при отображении 6 белых вертикальных полос на черном экране. Запись IQ сигналов выполнена с помощью анализатора RSA306B и программного обеспечения SignalVu-PC. На рис. 3 отображены значения корней из суммы квадратов IQ сигналов (мгновенные значения амплитуд). На фрагменте показаны 6 линий развертки монитора. В каждой линии просматриваются сигналы от 6 вертикальных полос. При получении записи VGA сигнала всегда возникает вопрос

синхронизации как по строкам, так и по кадрам. На рисунке 4 показан результат расчета автокорреляционной функции представленного фрагмента. Максимумы автокорреляционной функции позволяют обеспечить требуемую синхронизацию.

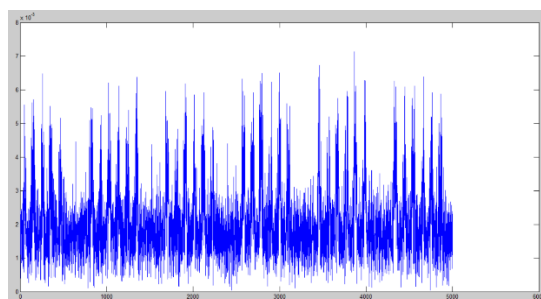


Рис. 3. Фрагмент записи 6 линий

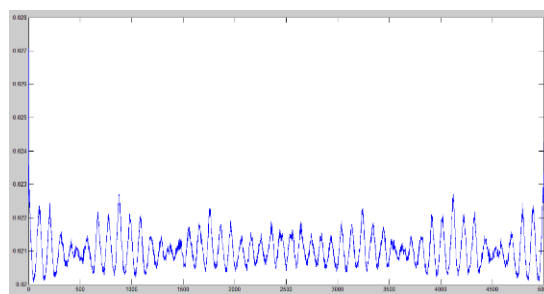


Рис. 4. Автокорреляционная функция

ВЫВОДЫ

Представленные результаты исследований подтверждают доступность формирования электромагнитного канала утечки информации на основе общедоступных радиоприемных устройств. Получена прямая регистрация изображения монитора с помощью телевизионного приемника, реализованного на основе анализатора спектра. Выполнена запись IQ сигналов ПЭМИ VGA интерфейса, открывающая новые потенциальные возможности по восстановлению изображений мониторов на фоне шумов.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Markus G. Kuhn*. Compromising emanations: eavesdropping risks of computer displays // Technical Report UCAM-CL-TR-577 University of Cambridge Computer Laboratory, 2003. С. 167.
2. *Martin Marinov*. Remote video eavesdropping using a software-defined radio platform // A dissertation submitted to the University of Cambridge in partial fulfilment of the requirements for the degree of Master of Philosophy in Advanced Computer Science, 2014. С. 68.
3. *Asanka Sayakkara, Nhien-An Le-Khac, Mark Scanlon*. Accuracy Enhancement of Electromagnetic Side-Channel Attacks on Computer Monitors.

АНАЛИЗ ШЕРОХОВАТОСТЕЙ ГЕТЕРОГРАНИЦ С ИСПОЛЬЗОВАНИЕМ ПЭМ ИЗОБРАЖЕНИЙ КВАНТОВО-КАСКАДНЫХ ГЕТЕРОСТРУКТУР

Ан. А. Афоненко¹, А. А. Афоненко¹, Д. В. Ушаков¹,
С.С. Пушкарев^{2,3}, Р.А. Хабибуллин^{2,3}

1) Белорусский государственный университет, Минск, Беларусь

E-mail: afonenko@bsu.by

2) Институт сверхвысокочастотной полупроводниковой электроники им. В.Г. Моке-
рова РАН, Москва, Россия

3) Московский физико-технический институт (национальный исследовательский
университет), Долгопрудный, Россия

В работе проведен анализ ПЭМ изображений квантово-каскадной гетероструктуры, содержащей 228 периодов из трех квантовых ям GaAs/Al_{0,18}Ga_{0,82}As и выращенной методом молекулярно-лучевой эпитаксии. Найдены размеры переходных областей. Получены спектры и автокорреляционные функции флуктуаций состава. Проведены оценки толщины и корреляционной длины шероховатостей гетерограниц.

Ключевые слова: квантово-каскадная гетероструктура; просвечивающая электронная микроскопия; шероховатости гетерограниц.

Квантово-каскадные лазеры (ККЛ), основанные на межподзонных переходах электронов в зоне проводимости гетероструктур GaAs/AlGaAs, являются уникальными компактными твердотельными источниками терагерцового (ТГц) излучения в диапазоне от 1,2 до 5,4 ТГц [1]. Важным этапом отработки технологии роста квантово-каскадных лазерных гетероструктур являются исследования, направленные на анализ структурного совершенства выращенных образцов [2, 3]. Сведения о толщинах переходных слоев и характеристиках шероховатостей границ необходимы для учета спектрального уширения энергетических уровней и корректного расчета спектров усиления и плотности тока через структуру.

Целью данной работы является анализ ПЭМ изображения (ПЭМ: просвечивающая электронная микроскопия) квантово-каскадной гетероструктуры Al_{0,18}Ga_{0,82}As/GaAs, выращенной методом молекулярно-лучевой эпитаксии, определение размеров переходных областей, оценка среднеквадратичных шероховатостей гетерограниц и их автокорреляционной функции.

ПРОФИЛЬ КОМПОНЕНТНОГО СОСТАВА ГЕТЕРОСТРУКТУРЫ

Для восстановления профиля гетерограниц в исследуемой гетероструктуре производилось суммирование интенсивности пикселей ПЭМ изображения (рис. 1) вдоль атомных плоскостей. При этом найденный наклон атомных плоскостей к горизонтали составлял 1,23–1,29°. Подгонка

восстанавливаемого профиля проводилась методом наименьших квадратов с использованием линейной интерполяции фона и нормальной функции размытия состава (рис. 2)

$$f_g(z) = \exp\left(-\frac{z^2}{2\tau^2}\right), \quad (1)$$

где τ – среднеквадратичная величина размытия.

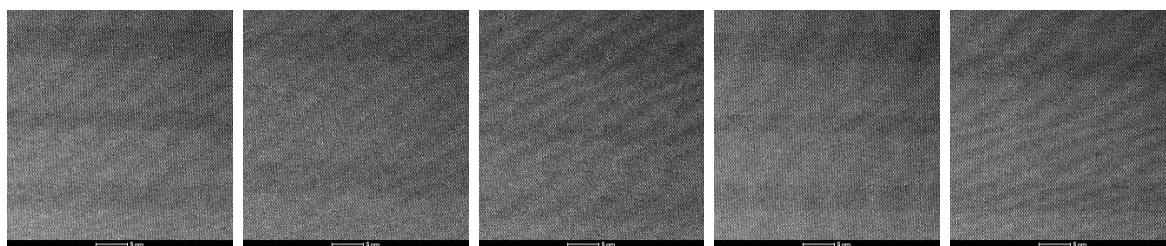


Рис. 1. ПЭМ изображения областей гетероструктуры квантово-каскадного лазера с разрешением 0,0352 нм на пиксель.

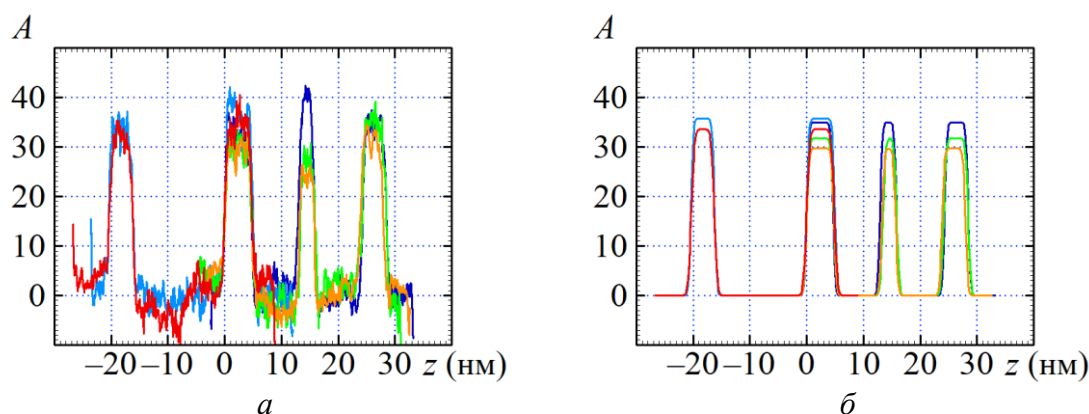


Рис. 2. Зависимость интенсивности пикселей A обработанных ПЭМ изображений от координаты z , направленной вдоль нормали к плоскости слоев (a) и восстановленный профиль состава (b). Начало координат для каждого профиля выбрано на границе широкой квантовой ямы и барьерного слоя

Результаты анализа приведены в таблице. Среднее размытие гетерограниц составило 0,37 нм, что в 2–3 раза меньше, чем для структур, выращенных методом МОС-гидридной эпитаксии [3]. Измеренная толщина 3-го слоя оказалась на 0,86 нм больше проектного значения. Для остальных слоев отклонение от проектных величин не превышает толщины монослоя 0,28 нм.

**Восстановленная амплитуда профиля A , размытие гетерограниц τ
и толщины слоев структуры d**

№ п.п.	A	τ , нм	d_1 , нм	d_2 , нм	d_3 , нм	d_4 , нм	d_5 , нм	d_6 , нм	d_7 , нм
1	34,9	0,34			4,65	8,31	2,89	8,48	4,15
2	35,7	0,34	4,15	16,22	4,96				
3	31,7	0,38			4,75	8,69	2,43	8,27	4,33
4	29,7	0,36			4,82	8,45	2,43	8,16	4,15
5	33,6	0,45	4,05	16,26	4,89				
Среднее	33,12	0,37	4,10	16,24	4,81	8,48	2,58	8,30	4,21
Проект			4,23	16,09	3,95	8,75	2,54	8,18	4,23

ФЛУКТУАЦИИ КОМПОНЕНТНОГО СОСТАВА ГЕТЕРОСТРУКТУРЫ

Для нахождения шероховатостей гетерограниц анализировались флуктуации состава в диапазонах $\pm\tau$ от установленных в предыдущем разделе гетерограниц. Спектральная плотность “мощности” флуктуаций получалась усреднением квадратов Фурье-компонент всех 26 гетерограниц (рис. 3 а). Резонансы полученной зависимости соответствуют периоду расположения атомов $\sim 0,4$ нм. Среднеквадратичное значение флуктуаций состава σ_x соответствует эквивалентной шероховатости гетерограницы

$$\sigma_d = 2\tau \frac{\sigma_x}{x_b}, \quad (2)$$

где x_b – средняя величина состава барьерного слоя.

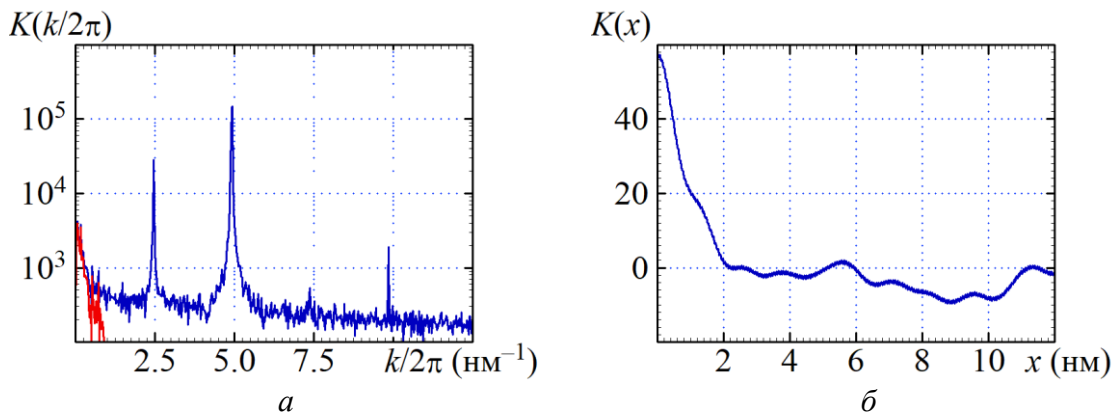


Рис. 3. Усреднённая по всем анализируемым гетерограницам спектральная плотность (а) и автокорреляционная функция флуктуации состава (б). Красным цветом показана спектральная плотность флуктуаций после вычитания белого шума и вклада высокочастотных резонансов, которая использована для получения автокорреляционной функции

Автокорреляционная функция флуктуаций получалась Фурье-преобразованием спектральной плотности. При этом учитывалась только низкочастотная часть спектра, из которой вычтен белый шум по его среднему уровню в высокочастотной области (рис. 3 б). Оцененное значение длины корреляции составило 1,0 нм, относительная среднеквадратичная флуктуация состава – 23 %, среднеквадратичная шероховатость гетерограниц – 0,17 нм. Следует отметить, что приведенные значения являются характеристиками изображения. Например, измеряемая среднеквадратичная величина флуктуаций $\sigma_{x\text{ТЕМ}}$ уменьшается обратно пропорционально квадратному корню из толщины образца L :

$$\sigma_{x\text{ТЕМ}} \approx \sigma_x \sqrt{\frac{2L_c}{L}}, \quad (3)$$

где L_c – длина корреляции. Поэтому для получения истинных параметров флуктуаций полученные значения необходимо корректировать [2].

Таким образом, в работе выполнен анализ профиля и шероховатостей гетерограниц на основе ПЭМ изображений гетероструктуры, выращенной методом молекулярно-лучевой эпитаксии. Найденные размеры переходных областей составили 0,37 нм для нормального закона размытия границ. Оцененная длина корреляции шероховатостей гетерограниц составила 1,0 нм, среднеквадратичное отклонение 0,17 нм. Полученная длина корреляции в несколько раз меньше величин, получаемых методом сканирующей электронной микроскопии. Это свидетельствует о существенном вкладе шероховатостей или загрязнений поверхности образца в флуктуации интенсивности анализируемого ПЭМ изображения.

Исследование выполнено при финансовой поддержке в рамках проектов БРФФИ Ф23РНФМ-064 и гранта Российского научного фонда № 24-49-10004, <https://rscf.ru/project/24-49-10004/>».

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Хабибуллин Р.А., Маремьянин К.В., Пономарев Д.С. Квантово-каскадный лазер на 3.3 ТГц на основе активного модуля из трех квантовых ям GaAs/AlGaAs с рабочей температурой >120 К. ФТП. 2021. Т. 55, вып. 11. С. 989–994.
2. Исследование интерфейсов гетероструктур квантово-каскадных лазеров ТГц диапазона / Афоненко Ан. А., Афоненко А.А., Ушаков Д.В. [и др.] // Труды XXV Международного симпозиума «Нанозифика и нанозлектроника». Том 2. Нижний Новгород: Изд-во Нижегородского госуниверситета, 2021. С. 566–567.
3. Анализ ПЭМ изображения квантово-каскадной лазерной гетероструктуры, выращенной методом МОС-гидридной эпитаксии. / Афоненко Ан. А., Афоненко А.А., Ушаков Д.В. [и др.] // Труды XXVII Международного симпозиума «Нанозифика и нанозлектроника». Том 2. Нижний Новгород: ИПФ РАН, 2024. С. 585–586.

ОПРЕДЕЛЕНИЕ ВЫСОТЫ ОРБИТЫ КОСМИЧЕСКОГО ОБЪЕКТА С ИСПОЛЬЗОВАНИЕМ ПАССИВНЫХ ОПТИЧЕСКИХ ИЗМЕРЕНИЙ

**В.С. Баранова, А.А. Спиридонов, В.Л. Козлов, Д.В. Ушаков,
В.А. Саечников**

*БГУ, Минск, Республика Беларусь
E-mail: vbaranova@bsu.by*

Представлен метод оценки высоты орбиты космического объекта с использованием динамических данных пассивной оптической системы наблюдения и геометрического соотношения между наклонной дальностью и скоростью смещение космического объекта в плоскости кадра.

Ключевые слова: космический объект; оптические измерения; высота орбиты.

Для оптических угловых измерений к классическим подходам решения задачи начального определения орбиты относятся различные аналитические и численные модели итерационного поиска параметра дальности. Наибольшее практическое значение из которых, нашли метод Лапласа, Гаусса, двойной итерации и метод Гудинга, а также их различные модификации [1]. Причем, каждый из этих методов ограничен условиями применимости. Основные ограничения накладываются на угловую и временную разницу между измерениями, а также на области распространения наиболее точного решения: межпланетные или околоземные космические объекты. В настоящее время развивается тенденция поиска универсальных методов начальной оценки орбиты, условия применения которых не будут ограничены ни количеством данных пассивных оптических измерений, ни значениями этих данных. Одним из таких направлений являются различные геометрические приближения.

Известно, что измерения пассивной оптической системой проводятся в топоцентрической экваториальной системе координат связанной с плоскостью местного горизонта. Геометрическое представление орбитального движения космического объекта относительно наземной станции наблюдения изображено на рис. 1. К основным величинам пространственного соотношения между дискретными точками плоскости орбиты космического объекта и плоскостью кадра точки наблюдения на интервале прямой видимости относятся: вектор наклонной дальности $\vec{\rho}$, высота орбиты H , угол между наклонной дальностью и направлением на подспутниковую точку θ , угол видимости φ , орбитальная скорость космического объекта нормальная к радиусу орбиты $\vec{V}$, проекция орбитальной скорости на плоскость кадра $\vec{V}_p$, угол места или элевация el , а также радиус Земли R_e .

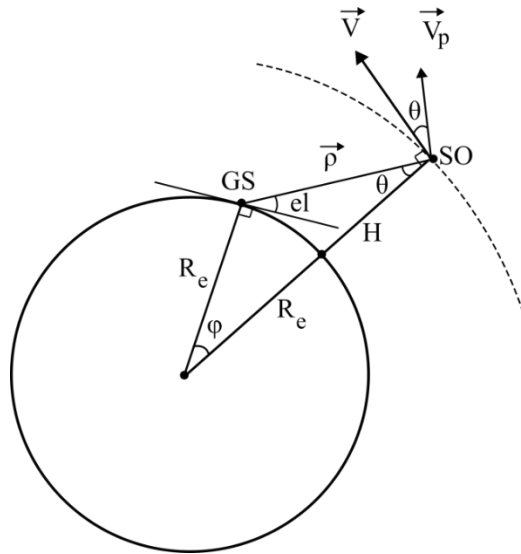


Рис. 1. Геометрия измерений пассивной оптической системой

Конечное нелинейное уравнение для расчета высоты орбиты космического объекта из соотношения наклонной дальности и скорости смещение космического объекта в плоскости кадра представляется как:

$$\frac{(R_e + H)^{3/2}}{\Delta t \sqrt{\mu}} \left[1 - \frac{R_e \sin(el)}{\left((R_e + H)^2 + (R_e \cos(el))^2 \right)^{1/2}} \right] = \frac{f}{\omega \Delta x},$$

где Δt время, за которое космический объект смещается в плоскости кадра на Δx пикселей, ω размер пикселя, f глубина фокуса, μ - гравитационный параметр.

Важно отметить, что именно точность определения глубины фокуса f будет существенно влиять на точность конечных расчетов высоты орбиты космического объекта. Глубина резкости будет изменяться в зависимости от величины наклонной дальности, а именно от того, отдаляется или приближается объект относительно точки наблюдения. Учитывая, что размер пикселя ω - величина постоянная, в дальнейшем глубина фокуса будет характеризоваться общим оптическим коэффициентом вида f / ω .

Для численного расчета высоты орбиты проводились серии измерений пассивной оптической системой наблюдения за низкоорбитальными космическими объектами Белорусского государственного университета [2]. Особенностью системы является использование видео данных и метода согласованной фильтрации для автоматического обнаружения космических объектов [3]. Исходные данные измерения для численного расчета высоты орбиты, согласно конечному нелинейному уравнению, представляют собой значения смещения спутника в плоскости кадра Δx за $\Delta t = 1$ с,

топоцентрические экваториальные координаты прямого восхождения α_t и склонения δ_t для выбранных точек, время измерения выбранных точек в UTC, оптический коэффициент f / ω для каждой точки измерения, а также геодезические координаты наземной станции наблюдения φ, λ . На рис. 2, а б представлена графическая зависимость оптического коэффициента f / ω от угла места el для данных угловых измерений спутников ADEOS-2 и AQUA соответственно.

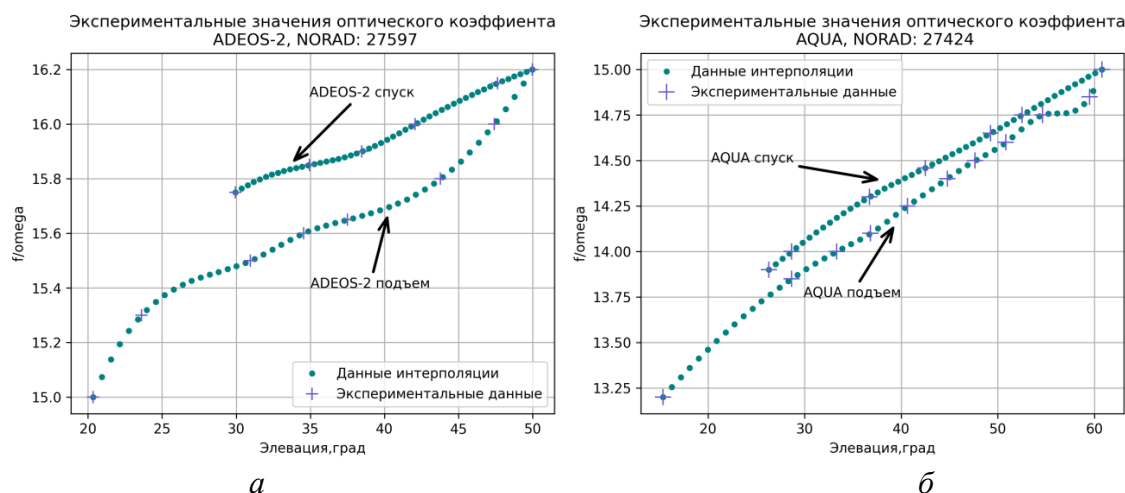


Рис. 2. Зависимость оптического коэффициента f / ω от угла места el

Как видно из графиков (рис. 2, а б) значение f / ω определяется величиной наклонной дальности ρ . При минимальном расстоянии между космическим объектом и наземной станцией наблюдения (при максимальном значении el) f / ω достигает своего максимального значения 16,2 для спутника ADEOS-2 и 15,0 для спутника AQUA. Разные значения скорости изменения и сдвига f / ω для ADEOS-2 и AQUA обусловлены разницей в высоте орбиты, которая для ADEOS-2 в среднем составляет $h=805$ км, а для AQUA- $h=704$ км.

Результаты расчета высоты орбиты спутников ADEOS-2 и AQUA согласно данным измерения оптической системы наблюдения и численному решению нелинейного уравнения представлены на рис. 3, а б. Анализ результатов показывает, что предложенный метод позволил определить высоту спутников ADEOS-2 и AQUA в пределах их апогейной(кривая с кружками рис. 3, а б) и перигейной высоты(кривая с треугольниками рис. 3, а б). В частности, ошибка в величине высоты для обоих спутников не превышает 4 км.

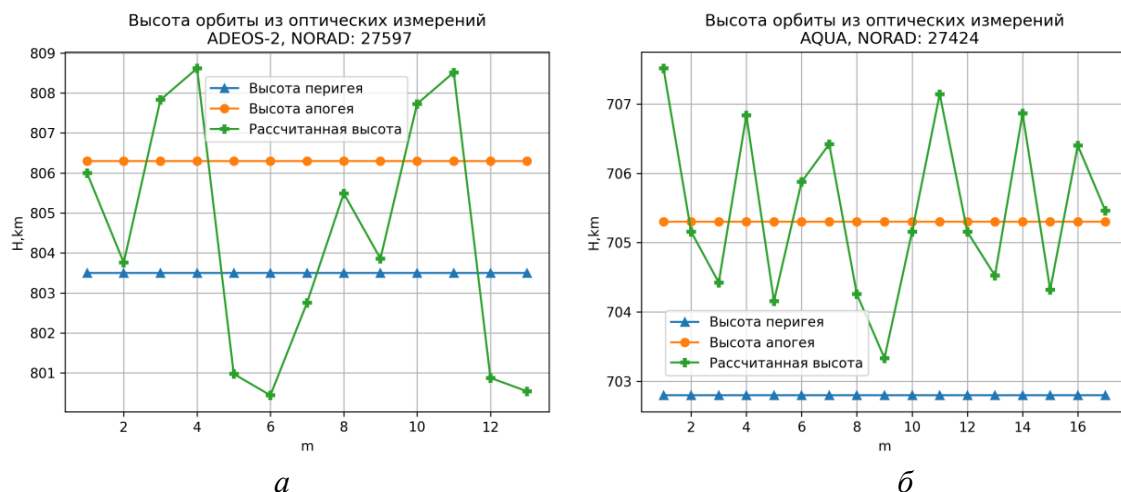


Рис. 3. Результаты расчета высоты орбиты

Как упоминалось ранее, точность определения оптического коэффициента f / ω критически влияет на точность расчета высоты. При точной настройке и автоматической записи глубины фокуса $\div$ до 3-го знака после запятой для каждой точки угловых измерений, возможно уменьшить ошибку до 1 км, что является предметом предстоящих исследований.

В работе представлены результаты расчета высоты орбиты спутников ADEOS-2 и AQUA согласно данным измерения оптической системы наблюдения и численному решению нелинейного уравнения, полученного из геометрического соотношения между наклонной дальностью и скоростью смещение космического объекта в плоскости кадра. Анализ полученных результатов показывает, что предложенный метод позволил определить высоту спутников в пределах их апогейной и перигейной высоты с ошибкой, не превышающей 4 км.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Vallado D. A. Fundamentals of Astrodynamics and Applications // Microcosm Press, 2013.
2. Baranova V., Spiridonov A., Liashkevich S., Saetchnikov V. Video Data Processing System for Ground-Based Space Optical Surveillance Application // 2023 IEEE 10th International Workshop on Metrology for AeroSpace (MetroAeroSpace), 2023.
3. Baranova V. S., Spiridonov A. A, Cherny V. A., Ushakov D. V., V. A. Saetchnikov. Resident space object detection method based on the connection between Fourier spectrum of the video data difference frame and the linear velocity projection // arXiv, 2024.

МЕТОДЫ ОЦЕНКИ ВЛИЯНИЯ КОСМИЧЕСКОЙ ПОГОДЫ НА БОРТОВЫЕ СИСТЕМЫ НИЗКООРБИТАЛЬНЫХ СПУТНИКОВ

Е.М. Глеба, В.С. Баранова, А.А. Спиридонов, В.Е. Черный, Д.В. Ушаков, В.А. Саечников

Белорусский государственный университет, Минск, Беларусь
E-mail: rct.gleba@bsu.by

В работе исследуется метод оценки влияния космической погоды на бортовые системы спутников на основе анализа телеметрии базы данных SatNOGS и выявления аномалий при помощи модели машинного обучения Polaris ML с использованием алгоритма XGBoost, реализующего метод Ньютона-Рафсона.

Ключевые слова: космическая погода; солнечная активность; аномалии телеметрии; SatNOGS; корреляционный анализ; модели искусственного интеллекта.

Введение. Космическая погода представляет собой комплекс явлений, происходящих в околоземном космическом пространстве, которые могут влиять на работу низкоорбитальных спутников. К ним относятся: солнечные вспышки, выбросы корональной массы, геомагнитные бури. Влияние космической погоды может проявляться в виде аномалий телеметрии, сбоев в работе бортовых систем и даже полной потери спутника [1]. В большинстве случаев, для обнаружения аномалий используются критерии на основе пороговых значений нижних и верхних пределов наиболее критичных параметров телеметрии бортовой электроники наноспутника. Также в последнее время, в область проектирования и эксплуатации наноспутников внедряются автоматизированные системы мониторинга состояния телеметрии, которые используют различные модели машинного обучения [2]. Подобные автоматизированные системы оценивают в целом работоспособность бортовых систем на момент испытаний или в режиме полетной диагностики. Спутник является сложной электрической системой, где события на каждом узле (компоненты бортовых систем) могут привести к последовательности не предугаданных сбоев. Причиной сбоев может послужить как естественные неполадки бортовых компонентов, так и внешние факторы. В данной работе проводится оценка взаимосвязи и влияния солнечной активности на работу бортовой электроники по данным телеметрии космических аппаратов.

Платформа Polaris ML. Приложение Polaris ML от LibreSpace использует алгоритм машинного обучения XGBoost для анализа взаимосвязей между параметрами телеметрии бортовых систем спутника. Polaris ML способен рассчитать и визуализировать взаимосвязи между параметрами телеметрии и степень их влияния друг на друга в виде

трехмерного графа связностей. В процессе исследования производительность модели искусственного интеллекта Polaris ML была значительно усовершенствована путем переноса core инфраструктуры на C++, а именно выделением прекомпилированных библиотек ввода/вывода и кэширования промежуточных расчетов обучения. Также в приложение Polaris были добавлены декодеры для отслеживания целевых спутников и модуль, собирающий информацию о солнечной активности, работающий в отдельном контейнере Docker. Кроме того, был доработан инструмент для построения графов связностей и веб-интерфейс. Внедрены решения для прямого взаимодействия с панелью управления сервера SatNOGS, что позволяет получать дополнительные данные о состоянии интересующих спутников.

На рис. 1 представлена архитектура финальной системы, включающая в себя внесенные доработки и усовершенствования.

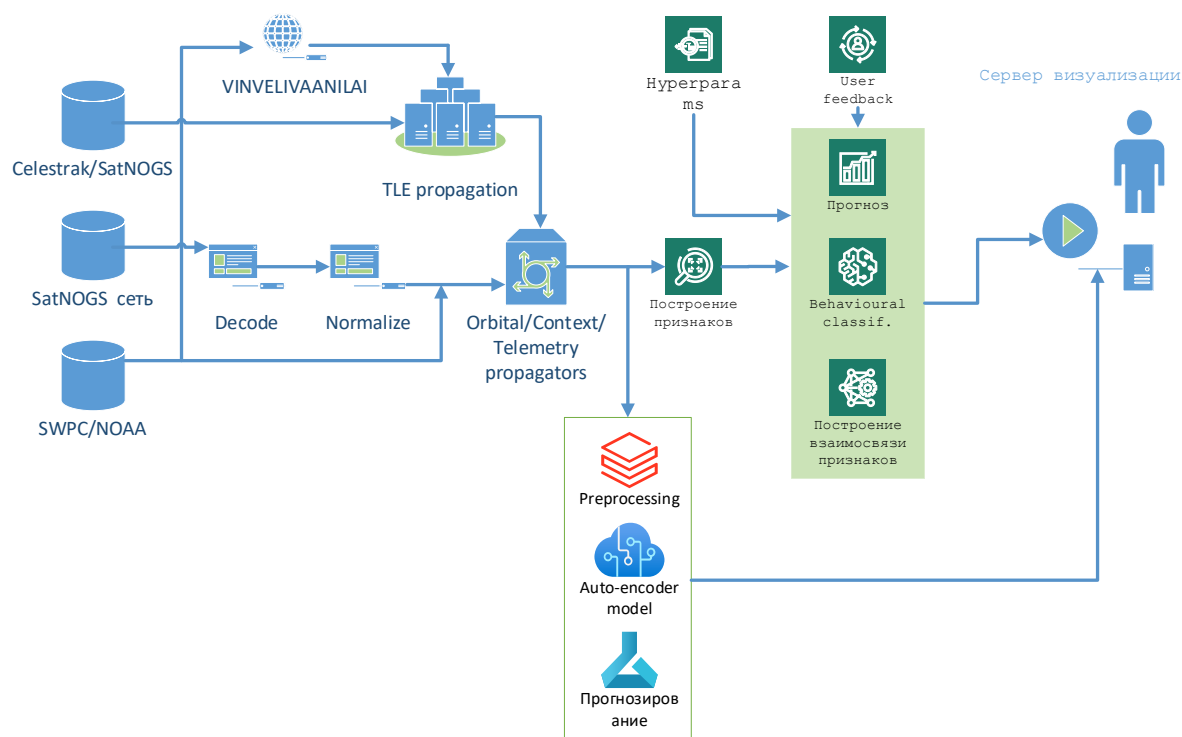


Рис. 1. Архитектура приложения Polaris

Этапы проведения анализа: 1) Автоматическое извлечение данных из различных источников. Включая данные телеметрии из сети SatNOGS и космической погоды NASA SWPC (NOAA); 2) Этап машинного обучения [4, 5] (XGBoost), на котором анализируется взаимосвязь всех извлеченных данных и формируется файл графика JSON в качестве выходных данных. Файл содержит описания параметров телеметрии и их значения в кадрах

с временными метками; 3) Визуализация построением графа связности, а также сложенного нормализованного вида телеметрии, отображающего аномалии параметров; 4) Экспортирование выходных данных графиков для дальнейшего анализа.

Как упоминалось ранее, был создан дополнительный модуль, который извлекает информацию о космической погоде с серверов NASA SWPC/NOAA. Эти данные сохраняются в базу данных InfluxDB, которая развернута локально в отдельном контейнере Docker Compose. Модуль также оснащен функциями для анализа данных орбитальных параметров в форматах TLE и OMM, а также для прогнозирования движения спутников на основе этих данных [6, 7].

Анализ перезагрузок основного процессор и граф связности. На рис. 2 (а) представлена взаимосвязь количества перезагрузок основного процессора наноспутника формата 3U GRIFEX от следующих индексов солнечной активности за период с 2019-03-03 по 2024-02-22: среднемесячные значения пятен S.I.D.C., SWPC/SWO и f10,7 см радиоизлучение. На рис. 2 (б) представлен результат использования модели Polaris ML для построения 3D графа связности параметров телеметрии спутника GRIFEX за период 2015–2021 г.

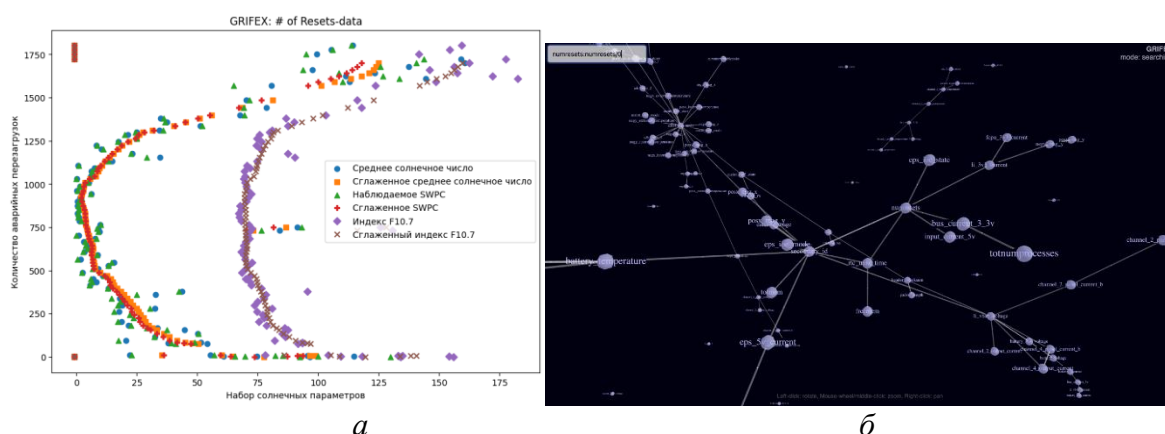


Рис. 2. а - диаграмма количества принудительных перезагрузок основного процессора с набором параметров активности Солнца для спутника GRIFEX; б - 3D граф связности параметров телеметрии спутника GRIFEX

Как видно, количество перезагрузок основного процессора GRIFEX стремительно накапливалось в период 2019–2020 г., когда индексы солнечной активности SWPC/SWO и f10,7 см были относительно постоянны. С 2020 начался новый 5-летний цикл солнечной активности, что соответствует замедлению роста количества перезагрузок. Причина такого поведения объясняется графом связности. Из анализа графа видно, все взаимо-

зависимые параметры телеметрии образуют взаимосвязанные ветки относительно времени (*rtc_unix_time*), а также относительно критичных параметров.

Общее количество перезагрузок основного процессора (*numresets*) непосредственно связано с токовым потреблением в бортовой системе питания, в частности, с значением тока на общей шине питания (*battery_bus_current*), с уровнем тока на шине 3,3 В (*bus_current_3_3v*), с уровнем тока на шине 3,3 В основного процессора (*fcpu_3v3_current*), с уровне тока на шине 3,3 В аккумулятора (*li_3v3_current*). Из графа также видно, что на количество перезагрузок процессора влияет число запущенных процессов (*totnumprocesses*). Можно предположить, что количество перезагрузок связано с просадкой напряжения на шине питания за счет низкой температуры на аккумуляторах (*battery_temperature*). При увеличении солнечной активности увеличился общий световой поток, что привело к повышению температуры, стабилизации напряжения и несвойственному замедлению роста числа перезагрузок процессора.

Таким образом, доработанная и адаптированная модель Polaris ML может быть использована для комплексного анализа взаимосвязей и влияния различных факторов на работу бортовой электроники наноспутников, в частности, солнечной активности. Оценка большего объема данных позволит выявить наиболее уязвимые системы спутника для разработки безопасных режимов работы его бортового оборудования в условиях агрессивной солнечной активности.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Green J. C., Likar J., Shprits Y.* Impact of space weather on the satellite industry // *Advancing Earth and space science*, 2017. № 15. С. 804–818.
2. *Schlag L., O'Meara C., Wickler M.* Numerical Analysis of Automated Anomaly Detection Algorithms for Satellite Telemetry // *SpaceOps Conference*, 2018. С.1-13.
3. XGBoost: A Scalable Tree Boosting System [Электронный ресурс]. URL <https://xgboost.readthedocs.io/en/latest/> (дата обращения: 07.03.2024).
4. *Boumghar R., Silva J., Angelis I., Schulster J., Donati A.* Enhanced awareness in space operations using multipurpose dynamic network analysis // *Space Operations: Inspiring Humankind's Future*. Springer International Publishing. 2018. P. 795-810
5. *Ray B.K., Tsay R.S.* Bayesian methods for changepoint detection in long-range dependent processes // *Journal of Time Series Analysis*. 2002. V. 23. P. 687–705.
6. *Bottou L.* Stochastic gradient learning in neural networks // *Proceedings of Neuro-Nimes*. 1991. V. 91.
7. *Killick R., Fearnhead P., Eckley I.A.* Optimal detection of changepoints with a linear computational cost // *J. Amer. Statist. Assoc.* 2012. V. 107. P. 1590–1598.

АВТОМАТИЧЕСКИЙ АНАЛИЗ ВЫЯВЛЕНИЯ ЗОН КОРНЯ НА ИЗОБРАЖЕНИЯХ ФЛУОРЕСЦЕНТНОЙ МИКРОСКОПИИ

Кабулов Т.А., Кезик С.В., Недзьведь О.В., Мацкевич В.С.,
Демидчик В.В.

Белорусский государственный университет, г. Минск;
kabulovtimur64@gmail.com

В работе рассматривается метод автоматического выделения зон корней *Arabidopsis thaliana* на изображениях флуоресцентной микроскопии.

Ключевые слова: флуоресцентная микроскопия; анализ изображений; сегментация; анализ корней; дистанционная карта.

ВВЕДЕНИЕ

В последние годы исследования в области феномики растений значительно продвинулись благодаря применению современных технологий визуализации. Одним из ключевых инструментов в этом направлении является флуоресцентная микроскопия, которая позволяет с высокой точностью наблюдать за структурами и процессами в живых организмах. Так, изучение структуры и функции корней растений с помощью флуоресцентной микроскопии открывает новые возможности для понимания их взаимодействия с окружающей средой и механизмов адаптации к различным условиям [1].

В корне выделяют несколько зон: зона корневого чехлика, зона деления, зона роста и растяжения, зона всасывания и проведения. Наблюдение за выделением активных форм кислорода (АФК) в различных корневых зонах имеет решающее значение, поскольку АФК играют важную роль в регулировании роста и развития растений, влияя на различные процессы в различных частях корня. Исследования показывают, что уровни АФК динамически генерируются или удаляются в растениях для адаптации к различным средам, что влияет на развитие корней. В зоне деления корня АФК необходимы для деления клеток, в то время как в зоне удлинения они способствуют дифференцировке клеток. Такое пространственное распределение АФК имеет важное значение для баланса между пролиферацией и дифференциацией корневых стволовых клеток как во время нормального развития растения, так и во время стресса [2].

Мониторинг АФК имеет решающее значение для определения конкретной роли АФК в реагировании на стресс и для разработки стратегий повышения устойчивости растений в неблагоприятных экологических условиях.

Сегментация изображений, полученных с помощью флуоресцентной микроскопии, является ключевым этапом анализа, позволяющим выделить интересующие объекты или зоны для дальнейшего изучения. В контексте корневых систем растений, сегментация зон корня на флуоресцентных изображениях представляет собой сложную задачу из-за высокой степени гетерогенности структур, перекрывающихся элементов и разнообразия форм. Автоматизация подобного процесса позволило бы сократить время анализа и улучшить его качество.

АЛГОРИТМ ВЫДЕЛЕНИЯ ЗОН КОРНЯ

Входными данными алгоритма являются изображения флуоресцентной микроскопии, которые приводятся к полутоновому виду (рис. 1а). Следующим шагом является бинаризация изображения корня (рис. 1б).

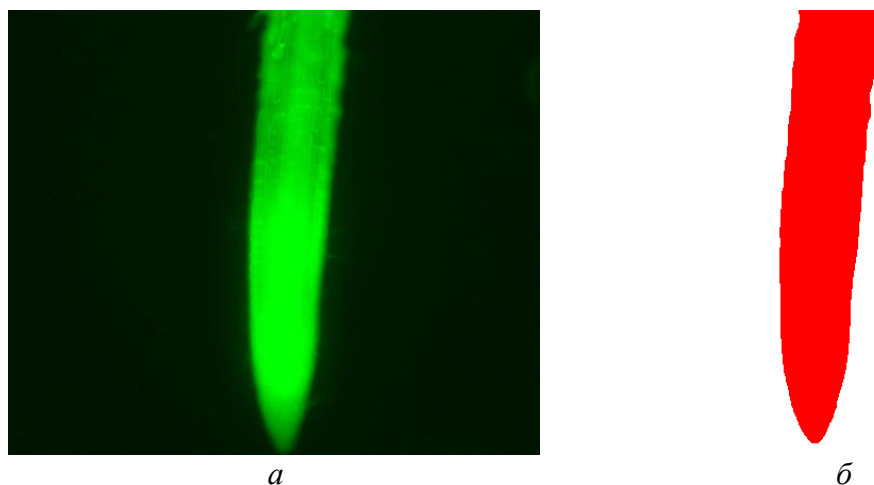


Рис. 1. Исходное изображение корня (а) и его бинарный образ (б)

Для бинарного изображения строится скелет, который отображает среднюю линию корня (рис. 2 а).

На следующем шаге строится дистанционная карта, которая отображает расстояние до края корня в каждой точке [3]. Это расстояние выражено в значениях пикселей на данной карте (рис 2 б).

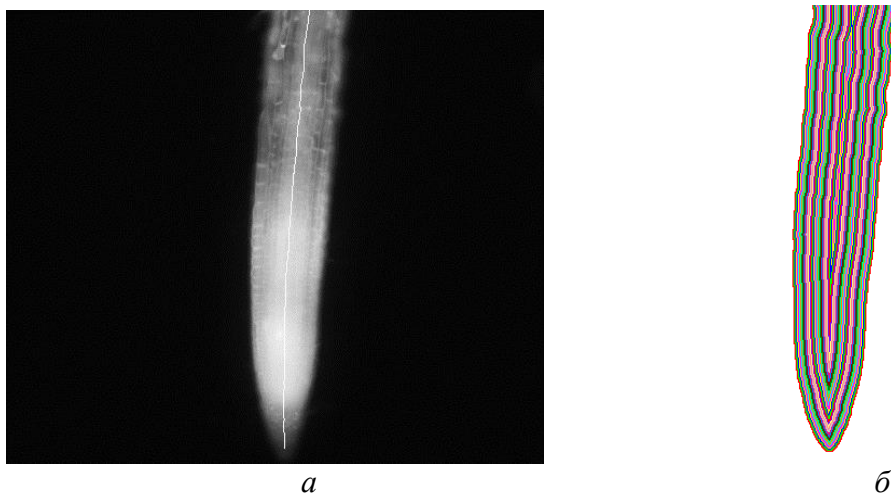


Рис. 2. Выделение средней линии на исходном изображении корня (а) и дистанционная карта (б)

При объединении средней линии с картой расстояний строится распределение ширины вдоль корня (рис. 3а). По этим значениям определяется коэффициент коррекции по ширине для каждого пикселя изображения вдоль корня (рис. 3б).

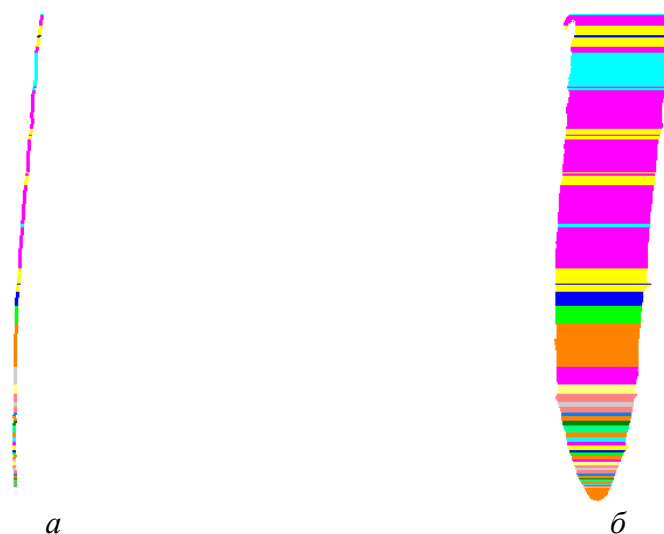


Рис. 3. Объединение дистанционной карты со средней линией (а) и распределение коэффициента ширины по образу корня (б)

Перемножение коэффициентов коррекции по ширине с яркостью исходных пикселей для области корня позволяет выделить зоны корня на изображении (рис. 4).

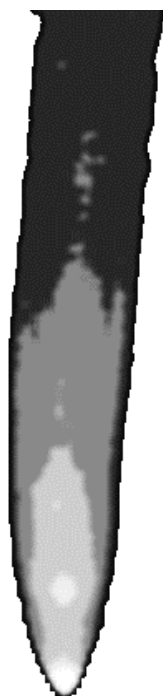


Рис. 4. Зона корневого чехлика, зона деления, зона роста и растяжения, зона всасывания и проведения выделены разными оттенками серого

ЗАКЛЮЧЕНИЕ

Данный алгоритм позволяет выделить зоны корня методами сегментации на изображениях флуоресцентной микроскопии. В дальнейшем это позволяет соотносить результаты мониторинга АФК с фазами роста и развития растения и служит для разработки стратегий повышения устойчивости растений в неблагоприятных экологических условиях.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *D. Wangenheim, R. Hauschild, J. Friml.* Light Sheet Fluorescence Microscopy of Plant Roots Growing on the Surface of a Gel // *Journal of Visualized Experiments*. 2017. Vol. 18 (119). doi: 10.3791/55044.
2. Mechanisms of ROS Regulation of Plant Development and Stress Responses/ *H. Huang [et al.]* // *Frontiers in Plant Science*. 2019. doi: 10.3389/fpls.2019.00800.
3. *Huang, X.* Two-Stage Segmentation Framework Based on Distance Transformation // *Sensors (Basel)*. 2021. Vol. 22(1). doi: 10.3390/s22010250

РАЗРАБОТКА ЛОКАЛЬНОГО АЛГОРИТМА ОГРАНИЧЕНИЯ ОБЛАСТИ ПОИСКА В ЗАДАЧЕ СТЕРЕОЗРЕНИЯ

К. Д. Колесников, Н. Н. Яцков, О. Ф. Ковалев

Белорусский государственный университет, Минск, Беларусь

E-mail: kolesnikovkirilld@gmail.com

В работе предложен локальный алгоритм ограничения области поиска. Процедура ограничения поиска является крайне необходимой в системах стереозрения – она позволяет избежать ложных соответствий. Особенно значительно проблема ложных соответствий проявляется на квазипериодических участках изображений. Работоспособность разработанного алгоритма подтверждается на примерах обработки наборов стереоизображений. Разработанный алгоритм улучшает качество получаемых результатов в среднем более, чем на 10%.

Ключевые слова: стереозрение; карта глубины; локальный алгоритм; оконное сопоставление стереоизображений; ограничение области поиска.

В последние годы довольно резко возрос интерес к системам автопилота и компьютерной графике. Фундаментальной задачей для этих направлений является задача определения расстояния до объектов сцены. Этой задачей занимается компьютерное стереозрение. Данные о расстоянии представляют собой матрицу, каждый элемент которой показывает расстояние до проецируемой на плоскость изображения точки сцены. Такую матрицу иначе называют картой глубины (диспаритета).

Для решения задачи стереозрения используются локальные и глобальные методы. Локальные строят карту инкрементально, глобальные – итеративно. Достоинствами локальных алгоритмов является быстродействие, а глобальных – точность. Локальные подразделяются на корреляционные – SAD, SSD, NCC, RT, CT (аббревиатуры обозначают метрику сходства областей или преобразование, SAD – сумма абсолютных отклонений, SSD – сумма квадратов отклонений, NCC – нормализованная взаимная корреляция, RT – ранговое преобразование, CT – преобразование переписи) и на методы, учитывающие особенности изображений, среди них – HCD (алгоритм детекции углов Харриса), SIFT (масштабно-инвариантное преобразование особенностей), SURF (ускоренный алгоритм надежного нахождения особенностей). В локальных методах не предусмотрен механизм работы с квазипериодическими областями.

Цель работы: разработать алгоритм локального ограничения области поиска, что уменьшит количество ложных соответствий, улучшить качество получаемых карт глубины (в сравнении с алгоритмом глобального ограничения области поиска), а также подтвердить его эффективность на практике.

МЕТОДОЛОГИЯ

На вход любой системы бинокулярного стереозрения подаются два согласованных изображения, полученных с разных ракурсов, – их называют стереопарой. Согласование изображений обычно происходит путем калибровки камер – они должны иметь параллельные оптические оси.

Построение качественной карты глубины представляется крайне нетривиальной задачей – это обусловлено рядом факторов. Качество карты снижают, к примеру, имеющиеся на изображениях отражения, окклюзии, неравномерность освещенности и т.п. Кроме того, существует проблема ложных соответствий – не получится гарантировать, что области были соотнесены корректно [1].

Важно заметить, что на реальных изображениях расстояние изменяется не плавно, а резко – в особенности при пересечении границ объекта. Логично разбить изображение на сегменты, охватывающие собой каждый конкретный объект сцены.

Подобная постановка задачи имеет сходство с задачей сегментации, однако алгоритмы сегментации, применяемые к решению данной задачи, не являются достаточно вычислительно эффективными [2]. В настоящей работе предлагается приближение способа сегментации, обладающее приемлемыми показателями как вычислительной эффективности, так и точности.

Предположим, что чем в более далеком сегменте изображения (относительно точки съемки) находится область, которой ищется соответствие, – тем сильнее должно быть ограничение области поиска.

Изложим алгоритм.

1. Вычислить корреляционную функцию между матрицами изображений стереопары.

Корреляционная функция вычисляется для сегментов пересечения изображений по линии их согласования (эпиполярной линии):

$$R_{m,n}(\tau) = \frac{\sum_{x,y \in S_{m,n}} (I_1(x,y) - \bar{I}_1)(I_2(x-\tau,y) - \bar{I}_2)}{\sqrt{\sum_{x,y \in S_{m,n}} (I_1(x,y) - \bar{I}_1)^2 \sum_{x,y \in S_{m,n}} (I_2(x-\tau,y) - \bar{I}_2)^2}}, \quad (1)$$

где I_1, I_2 – первое и второе изображение стереопары, $\tau \in [-N/8; N/8]$ – компенсация сдвига камеры, N – ширина изображения, $\bar{I}_1$ и $\bar{I}_2$ – средние значения интенсивностей пикселей на изображениях, а $S_{m,n}$ – сегмент пересечения $I_1(x,y) \cap I_2(x-\tau,y)$ с индексами m и n соответственно.

2. Ограничить область поиска внутри каждого сегмента по формуле:

$$MD_{m,n} = \frac{EC}{\max(R_{m,n}(\tau))} \cdot \operatorname{argmax}(R_{m,n}(\tau)), \quad (2)$$

где EC – некоторый весовой коэффициент.

3. Применить к полученным значениям $MD_{m,n}$ медианный фильтр 2-3 порядка, для сглаживания значений максимального отклонения.

4. Найти соответствия областей первого изображения на втором снимке. Для каждой области первого изображения вдоль эпиполярной линии (по горизонтали) определить наиболее схожую во втором, ограничивая область поиска рассчитанным максимальным отклонением $MD_{m,n}$, с индексами в соответствии с сегментом, которому область принадлежит.

5. Вычислить карту глубины. Рассчитать по парам соответствующих областей относительные расстояния, заполнить карту глубины. Применить к ней фильтры.

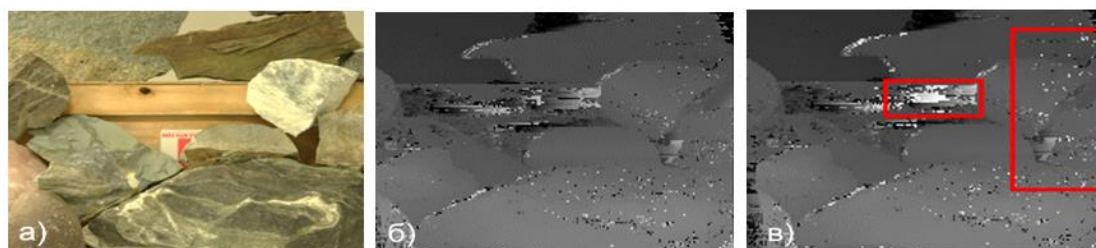
Можно заметить сходство предложенного алгоритма с алгоритмом, описанным в работе [3]. Принципиальным отличием является отдельная обработка сегментов изображения в задаче ограничения поиска – это обеспечивает локализацию алгоритмических ошибок.

Описанный метод предназначен в первую очередь для обработки стереоизображений с повторяющимися структурными элементами, в том числе полученных с систем, обладающих высокой разрешающей способностью по глубине, а также с преобладанием достаточно больших объектов (соразмерных используемым сегментам).

Вычислительная сложность. Пусть изображение стереопары имеет стороны размером N пикселей. Глобальный метод ограничения области поиска имеет асимптотическую сложность $O(N^3)$ [3]. Предложенный локальный метод имеет аналогичную сложность, однако значительно сокращает вычисления на шаге 4. Прирост производительности в среднем равен 10-20% (по сравнению с глобальным методом).

РЕЗУЛЬТАТЫ

В работе рассмотрены три стереопары с наличием квазипериодических областей (изображения 1080x720). Метод применен с параметрами: количество сегментов – 64, размер окна сканирования – 5, значения карты рассчитывались только для непересекающихся областей, метрика сходства – SAD. Пример результатов, полученных с помощью локального и глобального алгоритмов представлен на рисунке. Для оценки эффективности разработанного локального алгоритма (local search limiting method – LSLM) по сравнению с глобальным (global search limiting method – GSLM) рассмотрены критерии: визуальный анализ, время вычислений и оценка качества карты (результаты представлены в таблице). Оценка качества карты – максимум ее корреляционной функции R_M , на основе корреляции Пирсона, с истинной картой. Дополнительно исследовалось взаимодействие разработанного метода с методом адаптивного растяжения (adaptive stretch method – AStrM) [3].



а - Пример исходного изображения стереопары б - карта глубины, полученная методом LSLM, в - карта глубины, полученная методом GSLM

Сравнительная характеристика результатов работы методов

Метод		Пример 1		Пример 2		Пример 3	
		Время, с	R_M	Время, с	R_M	Время, с	R_M
Без AStrM	GSLM	14	0,46	20	0,52	22	0,05
	LSLM	11	0,53	14	0,81	17	0,09
Совместно с AStrM	GSLM	121	0,55	23	0,56	118	0,08
	LSLM	102	0,61	15	0,82	107	0,11

ЗАКЛЮЧЕНИЕ

В работе предложен локальный алгоритм ограничения области поиска LSLM. Предложенный алгоритм имеет следующие преимущества в сравнении с глобальным GSLM: позволяет лучше обрабатывать квазипериодические структуры на изображениях; увеличивает качество карты глубины (при небольшом размере окна сканирования более, чем на 10%) путем уменьшения количества ложных соответствий; во многих случаях отлично дополняет метод адаптивного растяжения AStrM.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Martin W. N., Aggarwal J. K. Motion Understanding: Robot and Human Vision // Springer Science & Business Media, 1988. ISBN 978-0-89838-258-7.
2. Альзаки Х. М., Цветков В. Ю. Текстурная сегментация изображений на основе оценки плотности контурных элементов и поглощения мелких областей // Минск, БГУИР, 2017, УДК 621. Р. 391.
3. Колесников К. Д., Яцков Н. Н., Ковалев О. Ф. Алгоритм определения сходства между однородными областями на стереоизображениях // Минск, НИИПФП, 2023, УДК 004.932.

РАЗРАБОТКА АВТОМАТИЧЕСКОГО ИНВЕСТИЦИОННОГО ПРИЛОЖЕНИЯ С ИСПОЛЬЗОВАНИЕМ ОН-ЧЕЙН МЕТРИК

Солодухо Н.А., Костенич А.А.

Белорусский государственный университет, Минск, Беларусь

E-mail: mikita.saladukha@gmail.com

Рассмотрены способы инвестирования и получения прибыли с использованием алгоритма торговли криптовалютами на ведущих мировых криптобиржах с использованием API (Application Programming Interface), предоставляемого криптобиржей ByBit. Разработанная программа использует он-чейн метрики для более чем 2000 торговых пар, что предоставляет широкий спектр инструментов для инвестирования. Результаты работы приложения показывают, что рынок криптовалют является очень молодым и быстро развивающимся, что придает ему большой показатель волатильности в отличие от традиционных рынков. Согласно полученной статистике работы приложения, мы можем убедиться в том, что институциональные инвесторы начинают вкладывать свои средства в рынок криптовалют. Таким образом приложение позволяет инвестировать вместе с крупными инвестиционными фондами.

Ключевые слова: криптобиржи; волатильность; традиционные рынки; инвестиционные фонды.

ВВЕДЕНИЕ

На сегодняшний день ряд крупных инвестиционных фондов продолжают совершать вложения в криптовалюты. Одним из таких является фонд BlackRock [1] — самый крупный и известный инвестиционный фонд в мире, во главе которого стоит Ларри Финк. Его компания на данный момент владеет 228 тысячами Биткоинами официально посредством ETF, а также миллионами Биткоинов посредством вложения в акции компаний по майнингу криптовалют, хедж-фонды, а также проводя раунды инвестирования в перспективные проекты.

Также стоит отметить, что с момента запуска Bitcoin ETF [2]. (11 января) компания Ларри Финка достигла состояния в 15 миллиардов \$, что показывает о перспективах развития криптовалют как с точки зрения инструментов экономики, так и для применения передовых технологий в финансовом секторе.

Целью данной работы является проверка гипотезы о возможности написания доступного, приносящего прибыль, полностью автоматического инвестиционного инструмента, который позволит инвестировать всем желающим без знаний и опыта в сфере инвестирования в перспективные криптовалютные проекты вместе с крупными фондами на примере BlackRock [1], при этом инвестирование в проекты упрощено до предела. Пользователям достаточно иметь аккаунт на бирже ByBit и пополнить счет, далее программа сама определяет проекты для инвестирования. В

основе алгоритма лежит пошаговое получение актуальных данных о проводимых операциях, объемах и изменении цены, их дальнейший анализ и принятие решения об инвестировании в ту или иную криптовалюту. Отличием данного приложения является абсолютная автономность, что не наблюдается во всех схожих существующих приложениях из области инвестирования в рынок криптовалют. Существует аналог на рынке Соединенных Штатов Америки, который называется Betterment [3], основывающийся на роботизированном инвестировании в Bitcoin ETF, что существенно отличается от разработанного приложения и ограничивается из всего списка криптовалют инвестированием в самую крупную по капитализации криптовалюту в мире — Bitcoin. Кроме того, существуют инвестиционные продукты у компаний Vanguard [4] и Fidelity [5], но, как и в случае с Betterment, ограничиваются инвестированием только в Bitcoin, и эти приложения все же требуют участия человека.

В основе инструментов для разработки приложения использовался язык программирования Python, так как это один из самых быстрых доступных языков в современном мире, допускающий взаимодействие с криптобиржей ByBit посредством API, нативно предоставляемом этой биржей [6]. При этом приложение не использует сторонних фреймворков или дополнительных библиотек, исключительно существующий базовый функционал предоставляемый языком Python.

ОПИСАНИЕ АЛГОРИТМА

Алгоритм представляет собой получение и дальнейший анализ торговых объемов, проводимых на криптовалютных биржах. На вход алгоритма идет json файл, содержащий информацию о всех поддерживаемых криптобиржах торговых парах. Далее получают необходимые данные для дальнейшего открытия торговой заявки: название торговой пары, её цену, объем за последние 24 часа. Далее производится расчет о том, как сильно изменился объем за секунду и открывает заявку на покупку данного актива. Критерием для открытия заявки на покупку актива является рост объема на активе в 100% (рис. 1). Рост в 100% был выбран из следующих соображений - рост объема в 100% является достаточным индикатором того, что в эту криптовалюту уже вложился крупный инвестор, что, скорее всего, будет способствовать и дальнейшему росту цены на криптовалюту. Благодаря быстрому API удастся с минимальной задержкой купить актив по рыночной стоимости до того, как он вырастет после обновления объемов на торговой бирже. При успешном выполнении одного цикла в результате происходит открытие торговой заявки на покупки и дальнейшее уведомление в Telegram бота об информации по совершенному торговой заявке.

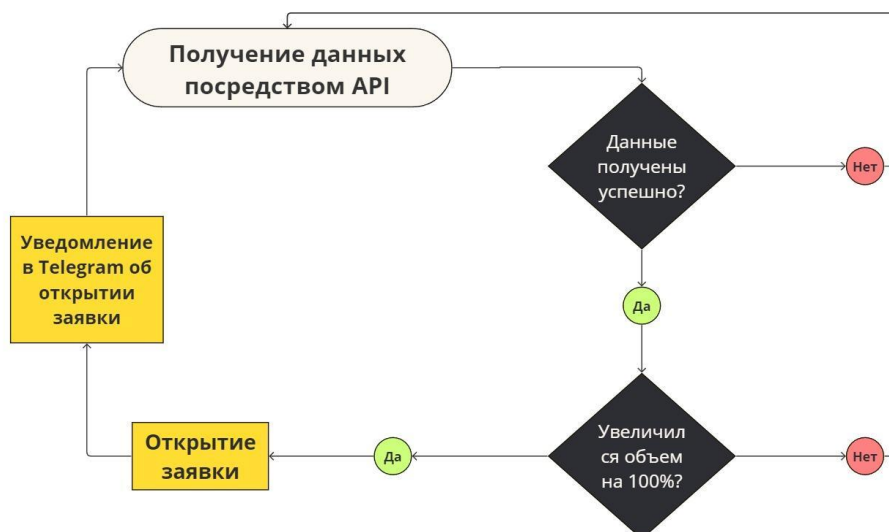


Рис. 1. Алгоритм работы приложения

Согласно полученной статистике работы приложения в виде торговых сделок за период в 4 неполных месяца было открыто 1067 торговых заявок только на бирже ByBit [3], что указывает на большую волатильность рынка и большом количестве торговых пар, в которые можно инвестировать. Наибольшее количество совершенных заявок на покупку было совершено в феврале, что указывает на перелив ликвидности из крупных криптовалют в криптовалюты с меньшей капитализацией. Кроме того, процент успешных сделок составляет 78,16%. Под успешными сделками понимается рост актива в дальнейшем на 50% от точки входа, а неудачными являются сделки с падением в 25% от точки входа. Такая модель инвестирования позволяет получать прибыль с коэффициентов $RR = 2$, так как каждая успешная сделка приносит прибыль в 2 раза больше, чем инвестор теряет в случае неудачи. Рост капитала под управлением приложения за период с декабря по март показан на рис. 2.



Рис. 2. Рост капитала под управлением приложения.

ЗАКЛЮЧЕНИЕ

В процессе разработки инвестиционного приложения были изучены основные подходы к инвестированию в рынок криптовалют, рассмотрены различные инструменты для взаимодействия с ведущими криптобиржами, а также реализовано полноценное автоматическое торговое приложение. Мы убедились, что гипотеза о возможности создания полностью автоматического, приносящего прибыль, инвестиционного приложения подтвердилась. Результаты показали, что рынок криптовалют только зарождается и в настоящий момент существует большое количество криптовалют, в которые может инвестировать любой, не имеющий серьезных познаний в данной области.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. BlackRock Foundation [Электронный ресурс]. URL: <https://www.blackrock.com/corporate/about-us> (дата обращения: 20.03.2024).
2. Statement on the Approval of Spot Bitcoin Exchange-Traded Products [Электронный ресурс]. URL: <https://www.sec.gov/news/statement/gensler-statement-spot-bitcoin-011023> (дата обращения: 20.03.2024).
3. Betterment Review 2024 [Электронный ресурс]. URL: <https://www.forbes.com/advisor/investing/robo-advisor-betterment-review> (дата обращения: 12.04.2024).
4. Vanguard [Электронный ресурс]. URL: <https://investor.vanguard.com/home> (дата обращения: 12.04.2024).
5. Fidelity [Электронный ресурс]. URL: <https://www.fidelity.com/managed-accounts/fidelity-go/overview> (дата обращения: 12.04.2024).
6. ByBit API [Электронный ресурс]. URL: <https://www.bybit.com/future-activity/en/developer> (дата обращения: 20.03.2024).

ИНТЕРФЕЙС КОМПЬЮТЕРНОЙ ПРОГРАММЫ «ПРОГНОЗИРОВАНИЕ ПСИХОНЕВРОЛОГИЧЕСКИХ ИСХОДОВ У НЕДОНОШЕННЫХ ДЕТЕЙ, РОДИВШИХСЯ С МАССОЙ ТЕЛА ДО 1500 ГРАММ»

Д.А. Крамко¹, В.С. Садов²

1) Республиканский научно-практический центр «Мать и дитя»

2) Белорусский государственный университет

г. Минск, Республика Беларусь

E-mail: kramko.dzmitry@gmail.com

В данной работе описан набор инструментов, который позволяет пользователю взаимодействовать с разработанной на основе искусственной нейронной сети компьютерной программой «Прогнозирование психоневрологических исходов у недоношенных детей, родившихся с массой тела до 1500 грамм», приведено визуальное представление программы, и набор функций, предоставляемых пользователю. Данная программа разработана для прогнозирования инвалидности у детей с пре- и перинатальным поражением центральной нервной системы, родившихся с экстремально низкой и очень низкой массой тела и, в зависимости от степени вероятности, для персонифицирования подхода к медицинской профилактике неблагоприятного исхода.

Ключевые слова: компьютерная программа; искусственная нейронная сеть; прогнозирование; инвалидность.

На основе математической модели прогнозирования инвалидности с применением технологии на основе искусственной нейронной сети разработана система прогнозирования неблагоприятных психоневрологических исходов у детей с пре- и перинатальным поражением центральной нервной системы, родившихся с экстремально низкой и очень низкой массой тела, для оказания медицинской помощи, представленная компьютерной программой «Прогнозирование психоневрологических исходов у недоношенных детей, родившихся с массой тела до 1500 грамм». Компьютерная программа разработана автором данной работы и зарегистрирована в реестре Национального центра интеллектуальной собственности [1, 2].

ОПИСАНИЕ КОМПЬЮТЕРНОЙ ПРОГРАММЫ

Компьютерная программа «Прогнозирование психоневрологических исходов у недоношенных детей, родившихся с массой тела до 1500 грамм» предназначена для врачей-педиатров, врачей-неврологов, врачей-реабилитологов, иных врачей-специалистов организаций здравоохранения II-IV технологических уровней, оказывающих медицинскую помощь детям.

Основным назначением данной компьютерной программы является прогнозирование неблагоприятных психоневрологических исходов (инва-

лидности) у детей с пре- и перинатальным поражением центральной нервной системы, родившихся с экстремально низкой и очень низкой массой тела (до 1500 грамм). Программа прогнозирует исход и указывает степень его вероятности. В зависимости от рассчитанной степени вероятности, проводится персонифицирование подхода к медицинской профилактике неблагоприятного исхода у ребёнка.

ФУНКЦИОНАЛЬНЫЕ ВОЗМОЖНОСТИ.

Вид основной экранной формы (визуального представления, формируемого на экране компьютера), сопровождающей процесс функционирования компьютерной программы, представлен на рис. 1.

Прогнозирование психоневрологических исходов у недоношенных детей, родившихся с массой тела до 1500 грамм

Настройки Справка

Пол	Мужской	Отмеченные факторы у пациента	Возрастной период
Вес при рождении, грамм	1150	Единственная артерия пуповины	Аntenатальный и интранатальный период
Срок гестации, недель	27	Обвитие пуповины вокруг шеи	Аntenатальный и интранатальный период
Рост при рождении, см	39	Ангиопатия сетчатки	Неонатальный период
Окружность головы, см	27	Врожденный порок развития ребер	Неонатальный период
Окружность груди, см	23	Врожденный порок сердца: дефект межпредсерд...	Неонатальный период
Оценка по шкале Апгар на 1-й минуте	6	Дисплазия тазобедренных суставов	Неонатальный период
Беременность по счету	3	РДС БГМ 3-4-ой степени	Неонатальный период
Роды по счету	2	Задержка речевого развития	Грудной возраст (до 1-го года)
От многоплодной беременности	Нет	Субкортикальные атрофические изменения	Результаты нейровизуализации

Возрастные периоды развития

- Аntenатальный и интранатальный период
- Неонатальный период
- Грудной возраст (до 1-го года)
- Результаты нейровизуализации

☐ Расширение межполушарной щели
☐ Расширение периваскулярных пространств белого вещества
☐ Расширение полости прозрачной перегородки
☐ Расширение субарахноидального пространства
☐ Ретроцеребеллярная киста
☐ Состояние после перенесенной гипоксии
☐ Состояние после перенесенной нейроинфекции
☐ Субатрофические изменения коры большого мозга
☒ Субкортикальные атрофические изменения
☐ Субependимальные кисты
☐ Туберозный склероз
☐ Шизэнцефалия

Поиск по названию:

Рассчитать прогноз

Прогноз **Благоприятный исход (инвалидности нет)**

Доверительная вероятность прогноза **52.01755%**

Рис. 1. Вид главного окна программы

Данная компьютерная программа создавалась в Microsoft Visual Studio на языке программирования C++ для наиболее распространённых операционных систем (Windows 7, 8, 8.1, 10, 11 со стандартными библиотеками), как традиционное классическое приложение, используя API Windows для отображения, с максимально простым интерфейсом, но защищённым от некорректного ввода данных.

Для работы с программой пользователю-врачу необходима информация о ребенке, полученная при рождении и в течение первого года жизни.

В левой верхней части главной экранной формы программы вводятся данные пациента при рождении: «Вес при рождении, грамм», «Срок гестации, недель», «Рост при рождении, см», «Окружность головы, см», «Окружность груди, см», «Оценка по шкале Апгар на 1-й минуте», «Беременность по счету», «Роды по счету» через элементы управления типа `NumericUpDown`, позволяющего как непосредственный ввод, так и выбор числа из определенного диапазона нажатием кнопки со стрелками вверх или вниз, либо прокруткой колёсика мыши. Данный тип элемента управления выбран для контроля ввода корректных данных параметров. Через элементы управления типа `ComboBox`, который позволяет ввод только одного значения из заранее предоставленного списка, вводятся пол пациента и каким по счёту он родился, если родился от многоплодной беременности.

Ниже находится элемент управления `ListView` – список возрастных периодов развития ("Аntenатальный и интранатальный период", "Неонатальный период", "Грудной возраст (до 1-го года)" и "Результаты нейровизуализации", при выборе одного из которых правее в элементе управления `CheckedListBox` отобразятся факторы (заболевания или особенности), характерные для выбранного возрастного периода развития недоношенных детей, родившихся с экстремально низкой и очень низкой массой тела, что позволяет наглядно отметить имеющиеся у пациента факторы.

В правой верхней части главного окна программы, в элементе управления `ListView` с двумя столбцами, отображается список отмеченных факторов пациента с указанием возрастного периода развития.

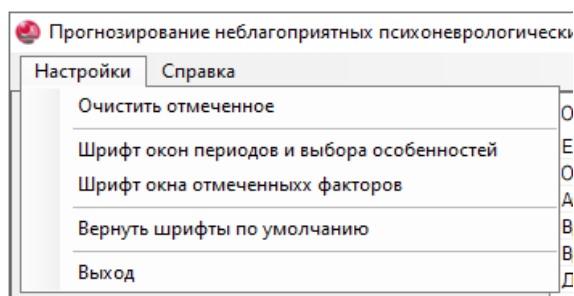


Рис. 2. Пункты меню «Настройки» с дополнительными функциями

В меню «Настройки», показанное на рис. 2, можно через пункт «Очистить отмеченное» отменить выбор всех отмеченных факторов у пациента; настроить шрифт отображаемых списков возрастных периодов и особенностей, настроить шрифт списка отмеченных факторов, вернуть

шрифты по умолчанию через одноименные пункты. В меню «Справка» содержится информация об авторах.

В нижнем левом углу главного окна программы расположена кнопка «Рассчитать прогноз». После ввода всех факторов и нажатия данной кнопки будет произведён расчёт прогноза психоневрологического исхода, и затем в нижнем правом углу в графе «Прогноз» отобразится результат: зеленая надпись «Благоприятный исход (инвалидности нет)», или красная «Неблагоприятный исход (инвалидность)» в зависимости от результата прогноза. Ниже будет отображена рассчитанная в процентах достоверительная вероятность данного прогноза.

На основе применения компьютерной программы «Прогнозирование психоневрологических исходов у недоношенных детей, родившихся с массой тела до 1500 грамм» разработана и утверждена Ученым Советом РНПЦ «Мать и дитя» инструкция по применению «Система прогнозирования психоневрологических исходов у детей с пре- и перинатальным поражением центральной нервной системы, родившихся с экстремально низкой и очень низкой массой тела», позволяющая определить вероятность развития инвалидности и в зависимости от степени вероятности персонифицировать подходы к медицинской профилактике неблагоприятного исхода. Производительность разработанной системы на тестовой выборке составила 89,29% [3]. Данная инструкция по применению внедрена в практику работы 11 организаций здравоохранения Республики Беларусь II-IV технологических уровней, оказывающих медицинскую помощь детям.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Девялтовская М.Г., Крамко Д.А., Винокурова И.А., Симченко А.В., Шумская Н.В. Психоневрологические исходы у недоношенных детей // Медицинские новости. 2019. № 5. С. 12-15.
2. Девялтовская М.Г., Крамко Д.А. Применение нейронных сетей для прогнозирования психоневрологических исходов у детей, родившихся с экстремально низкой и очень низкой массой тела // Сборник «Современные перинатальные медицинские технологии в решении проблем демографической безопасности». 2020. № 13. С. 310-315.
3. Крамко, Д.А. Медицинская профилактика инвалидности у детей, родившихся с экстремально низкой и очень низкой массой тела с применением технологий искусственного интеллекта // Первая выставка-форум IT-академграда «Искусственный интеллект в Беларуси»: сборник докладов, Минск, 13-14 октября 2022 г. Минск: ОИПИ НАН Беларуси, 2022. С. 114-119.

БОРТОВОЙ МОДУЛЬ ДЛЯ ТЕСТИРОВАНИЯ ИНЖЕНЕРНОЙ МОДЕЛИ СМКА

Д.В. Крупович, М.А. Павлышко, А.А. Камоцкий

Белорусский государственный университет, Минск, Беларусь

Email: krupovichdaniil@gmail.com, paulyshka@bsu.by, KamotskiAA@bsu.by

Предложен вариант построения модуля сверхмалого космического аппарата для тестирования систем инженерной модели университетского наноспутника. Описаны программно-аппаратные средства модуля системы диагностики. Описаны возможности разработанного программного обеспечения.

Ключевые слова: CAN; СМКА; ПО.

ВВЕДЕНИЕ

Космические аппараты, включая сверхмалые космические аппараты (СМКА, наноспутники), являются крайне сложными системами, требующими надежной передачи данных и управления. Одним из основных методов достижения надежной и эффективной передачи данных в подобных аппаратах является применение CAN-шины. В СМКА CAN-шина используется для организации обмена данными между различными системами управления, контроля энергообеспечения, системами безопасности, системой навигации и другими.

СМКА – это изделие, с которым невозможен близкий контакт человека-оператора и необходимы специальные средства связи для дистанционного управления, сбора экспериментальных данных, поступающих с него и диагностики систем аппарата [1]. Изучение проблем дистанционного контроля СМКА является одной из ключевых предпосылок к созданию системы диагностики инженерной модели СМКА, разработанного на факультете радиофизики и компьютерных технологий БГУ.

Данная модель сверхмалого космического аппарата разработана по принципу «слотовой» компоновки элементов. По габаритам и компоновке соответствует стандарту Cubesat 3U. Габариты корпуса позволяют использовать ИМ СМКА в условиях учебных лабораторий. Слотовая компоновка предоставляет возможности для расширения функциональности СМКА и комбинирования различных функциональных модулей для обработки различных исследовательских задач.

Модель СМКА поддерживает следующие интерфейсы передачи данных: SPI, I2C, USB, UART, CAN. Модули подключенные к материнской плате СМКА обмениваются данными между собой используя собственный протокол на базе CAN-шины.

Для связи с оператором, инженерная модель использует радиоканал с несущей частотой 436 МГц [5]. Однако, использование радиоканала на

регулярной основе затруднено: требуется наличие сложного аппаратно-программного комплекса для передачи данных. В связи с этим фактором было решено разработать бортовой модуль, который не будет требовать значительных затрат, а также будет являться более эффективным по своим характеристикам.

Сравнение характеристик каналов связи

Параметр сравнения	ESP32-C6-WROOM-1	E10-433MD-SMA
Протокол связи	Bluetooth 5.2	AX.25 G3RUH GFSK
Скорость передачи данных	125 Кбит/с, 500 Кбит/с, 1 Мбит/с, М Кбит/с	1,2 - 1000 Кбит/с
Рабочая частота кристалла	40 МГц	26 МГц
Дальность связи	до 150 м	до 2 км

ESP32-C6-WROOM-1 обладает более высокой пропускной способностью и скоростью передачи данных благодаря использованию технологии Bluetooth 5, что позволяет обеспечить более быструю и эффективную передачу данных. ESP32-C6-WROOM-1 обеспечивает более надежное соединение благодаря использованию передовых технологий и протоколов передачи данных.

АППАРАТНОЕ ОБЕСПЕЧЕНИЕ

Для облегчения коммуникации с СМКА был разработан модуль дистанционного контроля систем СМКА. Разработанный модуль включает в себя следующие компоненты (рис. 1):

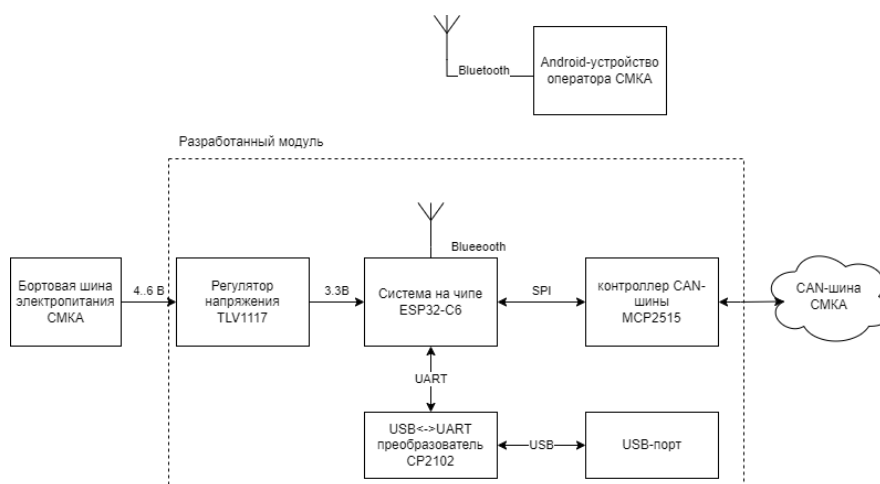


Рис. 1. Схема подключения разработанного диагностического модуля

1. Система на чипе Espressif ESP32-C6-WROOM-1. состоит из высокопроизводительного (HP) 32-битного процессора RISC-V с тактовой частотой до 160 МГц, встроенного радиомодуля Wi-Fi 6 (802.11ax) 2,4 ГГц, Bluetooth 5 (LE) и радиосвязь IEEE 802.15.4.

2. преобразователь USB-UART Silicon Labs CP2102. Данная микросхема используется для обеспечения возможности прошивки системы на чипе ESP32-C6-WROOM-1 посредством интерфейса USB..

3. CAN-контроллер MCP2515 для подключения системы на чипе ESP32-C6 к CAN-шине СМКА.

4. Дополнительные компоненты для мониторинга состояния модуля и обеспечения стабильного энергоснабжения модуля.

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ

Для управления разработанным модулем было создано программное обеспечение (ПО) для микроконтроллера ESP32-C6-WROOM-1. Оно предоставляет широкий спектр функций, позволяющих передавать CAN-сообщения в шину, получать их, осуществлять фильтрацию и пересылку на устройство оператора сверхмалого космического аппарата (СМКА) посредством Bluetooth. Таким образом, ПО обеспечивает полноценное управление и мониторинг работы модуля, включая взаимодействие с CAN-шиной и передачу данных на мобильное устройство оператора для дальнейшего анализа и управления.

Также было разработано ПО на базе операционной системы Android для мобильного устройства, которое связывается с диагностическим модулем и управляет им. Данные из инженерной модели СМКА передаются на мобильное устройство по Bluetooth. На стороне мобильного приложения пользователю предоставляется множество возможностей в области диагностики и тестирования. Имеющийся функционал мобильного ПО позволяет не только следить за обменом данных, но и активно управлять этим процессом и вмешиваться в него, делая тестирование более эффективным и интерактивным.

Приложение способно отображать информацию о статусе модуля: его текущем состоянии (включен или выключен), значениях текущего напряжения и тока (рис. 2). Кроме того, цвет фона ячейки может меняться для наглядной сигнализации об опасности: например, при превышении значений тока или напряжения относительно установленных норм. Это позволяет оперативно обнаруживать и реагировать на потенциальные проблемы, связанные с энергетическими модулями, упрощая процессы мониторинга и предотвращения аварийных ситуаций. Кроме того, приложение способно управлять системой энергоснабжения спутника, позволяя включать и выключать различные подключенные к нему модули.

ОПЕРАТИВНОЕ ТЕСТИРОВАНИЕ МИКРОСХЕМ АЦП

**А.В. Лебедевский, В.А. Фираго, Н.В. Левкович,
К.И. Шулико, В.М. Слодинская**

ФРФикТ БГУ, Минск, Беларусь

E-mail: firago@bsu.by

Представлено устройство на основе микроконтроллера для оперативного тестирования статической передаточной характеристики микросхем АЦП поразрядного уравнивания. Описана схема тестирования и приведены полученные результаты для трех типов АЦП, приобретенных малыми партиями. Результаты показывают наличие отклонения реальных параметров протестированных АЦП от заявленных в документации.

Ключевые слова: микропроцессорные системы; АЦП; тестирование; ШИМ.

ВВЕДЕНИЕ

Многим разработчикам при регистрации аналоговых сигналов приходится сталкиваться с проблемой выбора микросхемы АЦП с необходимыми параметрами и ее последующего тестирования. В документации на микросхемы АЦП приводится полный перечень их параметров, в том числе и максимальная частота выборок. Из них обобщающей характеристикой, описывающей основные свойства конкретной микросхемы АЦП, является статическая передаточная характеристика. Она позволяет оценить основные параметры АЦП: дифференциальную и интегральную нелинейности преобразования. В документации на широко используемые 12-ти и 16-ти разрядные микросхемы АЦП последовательного приближения дифференциальная нелинейность обычно меньше младшего уровня преобразования, а интегральная – не более одного-двух уровней.

При приобретении небольших партий АЦП приходится сталкиваться с несоответствием параметров АЦП, заявляемым в документации, поскольку частные продавцы не тестируют поставляемые компоненты. Поэтому актуально иметь простое микропроцессорное устройство, позволяющее проводить оценку соответствия измеряемой статистической характеристики тестируемой микросхемы АЦП параметрам, приводимым в ее документации.

УСТРОЙСТВО ТЕСТИРОВАНИЯ НА МИКРОКОНТРОЛЛЕРЕ STM32F103C8T6

При оценке интегральной нелинейности АЦП приходится сталкиваться с проблемой формирования пилообразного сигнала с высокой линейностью. Применение микросхем ЦАП в этом случае нежелательно, поскольку они имеют свою нелинейность преобразования. Поэтому целесообразно формировать нарастающий и спадающий фронты тестирующего

сигнала с помощью ЦАП на основе схемы широтно-импульсной модуляции (ШИМ), имеющейся в большинстве микроконтроллеров.

Подходящей для создания микропроцессорного устройства тестирования АЦП является плата отладки Blue Pill, использующая микроконтроллер STM32F103C8T6 от компании STM32. Тактовая частота используемого в ней микропроцессора составляет $f_c = 72$ МГц, что позволяет формировать с помощью ШИМ 14-разрядный трапециевидный тестирующий сигнал с высокой линейностью фронтов нарастания и спада за приемлемый интервал времени. Формируя с помощью 14-ти разрядной ШИМ трапециевидный сигнал, можно обеспечить длительность периодов фронтов нарастания или спада, равную $\tau_\phi = 2^{28} / f_c \approx 3,73$ с.

Недостатком такого решения является необходимость применения четырехзвенного низкочастотного RC фильтра. Расчеты показывают что при использовании номиналов $R = 62$ кОм и $C = 10$ нФ, граничная частота этого фильтра составляет 30 Гц, ослабление на частоте формирования ШИМ $f_{\text{ШИМ}} = 4394,53125$ Гц примерно 99 дБ, а фазовый сдвиг на частоте 0,26822 Гц, соответствующей периоду формирования пилю, около $0,6^\circ$. Максимальная амплитуда первой гармоники с частотой $f_{\text{ШИМ}}$, которая формируется на выходе этого фильтра около 25 мкВ, что достаточно для оперативного тестирования микросхем АЦП. Но возникают небольшие задержки фронтов нарастания и спада.

Для снижения времени тестирования АЦП частоту формирования одного шага ШИМ-сигнала необходимо выдерживать максимально возможной, т. е. до тактовой частоты микроконтроллера f_c . В таймерах микроконтроллеров STM32 реализована аппаратная поддержка генерации ШИМ-сигнала, используя которую можно генерировать трапециевидный сигнал.

Для обеспечения непрерывного измерения и непрерывной передачи данных используется двойная буферизация, при которой на микроконтроллере используются два буфера, в один из которых информация помещается, а другой в это время передается на ПК. При завершении записи в первый буфер их роли меняются местами и процесс повторяется заново. Поскольку время изменения их ролей незначительно (порядка нескольких сотен наносекунд), невынужденных задержек не наблюдается.

Для контроля стабильности работы программы, вместе с отправляемыми кодами АЦП на верхний уровень передается интервал времени (с микросекундной точностью [1]), потребовавшегося на заполнение буфера. Используя эти интервалы, можно измерять частоту съема данных. Экспериментально установлено, что микроконтроллер выдерживает заданную частоту $f_c / 2^{14} = 4394,53125$ Гц с неопределенностью менее 10 мГц.

РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ МИКРОСХЕМ АЦП

Исследовались три типа АЦП последовательного приближения. Были выбраны 12-разрядный МСР3201-С и 16-ти разрядный AD7683 с частотой выборок до 100 кГц. В качестве третьего образца использовался 12-разрядный АЦП, встроенный в микропроцессор STM32F103C8T6. Все они запитывались от стабилизированного источника напряжения +3,3 В. Поэтому цена младшего разряда преобразования составила около 0,8 мВ для 12-ти разрядных и 50 мкВ для 16-ти разрядных АЦП.

Для снижения влияния помех рекомендуется использовать дифференциальный вход АЦП, что усложняет схему. Поэтому с целью тестирования в условиях, приближенных к реально используемым на практике, сформированный трапециевидный сигнал подавался на неинвертирующий вход АЦП, а инвертирующий заземлялся. Никаких специальных мер по снижению наводок от работающей платы Blue Pill не предпринималось. Частота выборок АЦП была согласована с частотой формирования ШИМ $f_c / 2^{14}$, поэтому требования теоремы Котельникова не выполнялись, что приводит к возрастанию влияния на результаты исследований помех и шума. Особенно это заметно, как следует из рис. 1, при динамическом изменении входного сигнала. Шумовая компонента растет по мере возрастания входного напряжения. Но для МСР3201-С дифференциальная нелинейность ΔD_d не превышает 1-2 уровней преобразования АЦП. У встроенного АЦП ΔD_d гораздо хуже (около 20 уровней) из-за тактовых наводок. Сильное влияние помех на ΔD_d (около 60 уровней) наблюдается и для AD7683, что указывает на необходимость обязательного использования для 16-ти разрядных АЦП дифференциального входа.

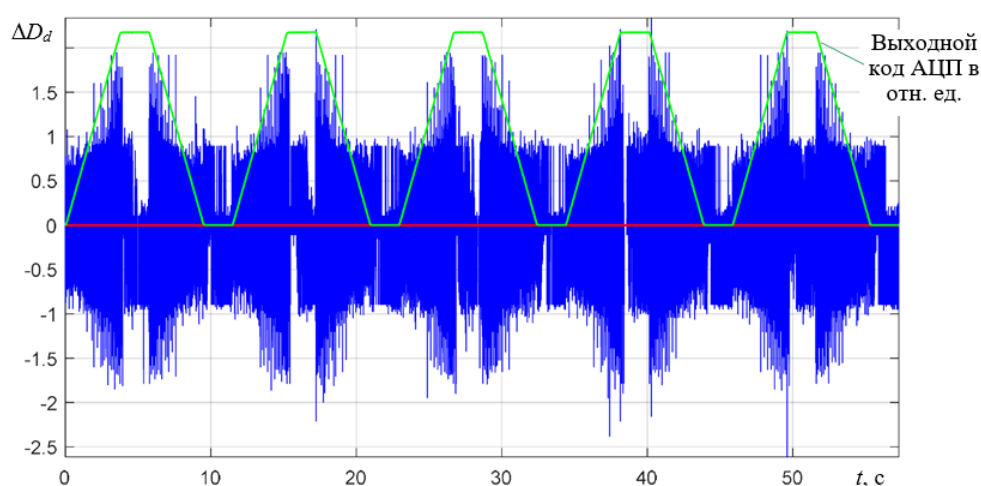


Рис. 1. Результаты исследования дифференциальной нелинейности АЦП МСР3201-С для 5-ти периодов трапециевидного входного сигнала

Интегральную нелинейность $\Delta D_I(D)$ определять сложнее, поскольку наблюдаемая задержка формирования трапецевидного сигнала из-за фазового сдвига применяемого НЧ фильтра вносит существенное смещение в получаемые на начальном интервале фронта (0–300) значения ΔD_I . Если их отбрасывать, полученные для МСР3201-С значения ΔD_I , как видно из рис. 2, примерно совпадают с заявленными в документации.

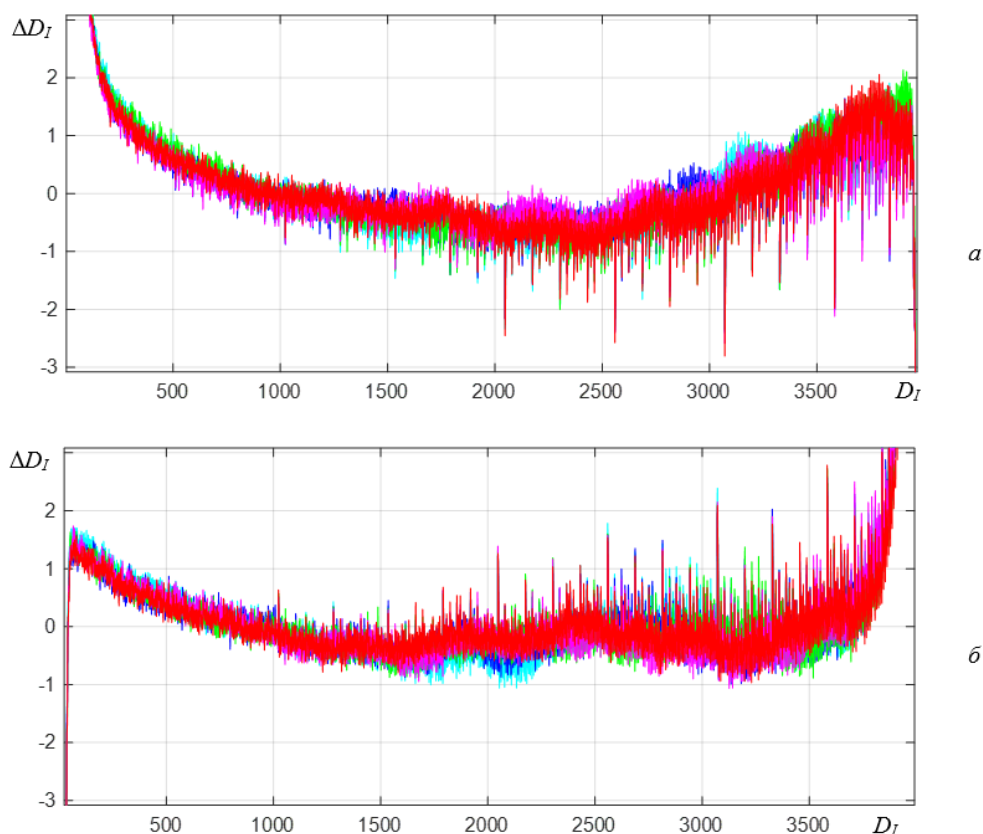


Рис. 2. Результаты исследования интегральной нелинейности МСР3201-С при нарастающем (а) и падающем (б) фронтах трапецевидного сигнала

Отклонения $\Delta D_I(D)$ для АЦП, встроенного в STM32, составила около 5 уровней преобразования, что меньше влияния помеховой компоненты (около 36 уровней), создаваемой тактовыми наводками работающей платы Blue Pill. При применяемой схеме включения AD7683 получаемые зависимости $\Delta D_I(D)$ гораздо хуже (максимальное отклонение достигает 40 уровней), чем заявленные в документации.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Cortex-M3 Technical Reference Manual r2p0 [Электронный ресурс]. URL: <https://developer.arm.com/documentation/ddi0337/h/data-watchpoint-and-trace-unit/about-the-dwt> (дата обращения: 30.03.2024).

ДВУХВОЛНОВАЯ РЕФРАКЦИОННАЯ ПУЛЬСОКСИМЕТРИЯ

В.А. Фираго, Н.В. Левкович, А.В. Лебедевский,
С. Г. Славинский, К.И. Шулико

ФРФУКТ БГУ, Минск, Беларусь

E-mail: firago@bsu.by

Установлено, что применение формул двухволновой трансмиссионной пульсоксиметрии при использовании модуля пульсоксиметра МАХ30102 приводит к значительному смещению измеряемых значений сатурации SaO_2 гемоглобина артериальной крови кислородом. Показано, что учет оптических параметров биоткани позволяет выявить пути коррекции расчетных выражений SaO_2 в рефракционной пульсоксиметрии.

Ключевые слова: сатурация гемоглобина; рефракционная пульсоксиметрия; формулы пульсоксиметрии; модуль пульсоксиметра.

ВВЕДЕНИЕ

В медицине применяют два способа двухволновой пульсоксиметрии [1]. В первом трансмиссионном регистрируют потоки излучения двух светодиодов (красного цвета свечения R и инфракрасного IR), прошедшие через контролируемый участок тела (мочка уха, палец руки). Поскольку малые артериальные сосуды микроциркуляторного русла пульсируют, возникает амплитудная модуляция прошедшего излучения. Определяя размах пульсаций регистрируемых сигналов ΔI и их минимальное значение $I_{\min}$ берут их отношение и далее на основе спектров поглощения оксигенированного HbO_2 и неоксигенированного Hb гемоглобина рассчитывают сатурацию артериальной крови SaO_2 при заданных значениях двух длин волн λ_1 и λ_2 , которые обычно соответствуют максимумам спектра излучения светодиодов (обычно 660 и 940 нм).

Применение модуля пульсоксиметра МАХ30102 [2, 3] позволяет минимизировать размеры датчика, поэтому его встраивают в различные устройства вплоть до браслета для фитнеса, позволяющего наряду с частотой пульса отслеживать и текущие значения SaO_2 . Его миниатюрная конструкция позволяет регистрировать излучение светодиодов (с максимумами спектра излучения на длинах волн 660 и 880 нм) которые рассеиваются кожей и подкожной тканью и попадают на рядом расположенный кремниевый фотоприемник. МАХ30102 относят к классу рефракционных пульсоксиметров, что требует более тщательного учета оптических характеристик ткани по сравнению с трансмиссионной пульсоксиметрией. Действительно проведенные нами измерения SaO_2 на дистальной фаланге среднего пальца руки группы людей при применении МАХ30102 и формул трансмиссионной пульсоксиметрии показывают существенное зани-

жение получаемых значений. В документации производителя [3] приводится аппроксимационная формула для расчета значений $SaO_2 = 104 - 17R$, где $R = (\Delta I_R / I_{Rmin}) / (\Delta I_{IR} / I_{IRmin})$, при использовании которой показания на 1–2% ниже, чем образцового пульсоксиметра. Формула для определения SaO_2 в фитнес браслетах [2] на основе MAX30102, которые располагаются на запястье руки, имеет другой вид $SaO_2 = 108,1 + 5,9\gamma - 44,6\gamma^2$, где $\gamma = \ln(\Delta I_{IR} / I_{IRmin}) / \ln((\Delta I_R / I_{Rmin}))$. Причем как в документации [2], так и в исследовании [3] не поясняется каким образом были получены применяемые выражения. Эти факты указывают на необходимость рассмотрения влияния оптических свойств биоткани на вид выражения, используемого для расчета SaO_2 .

ФОРМУЛЫ ТРАНСМИССИОННОЙ И РЕФРАКЦИОННОЙ ПУЛЬСОКСИМЕТРИИ

При их выводе используют спектральные (молярные или натуральные) показатели поглощения HbO_2 и Hb , которые приведены на рис. 1. В видимой области спектра значения $\mu_a^{HbO_2}(\lambda)$ и $\mu_a^{Hb}(\lambda)$, скорректированные с учетом фактора упаковки гемоглобина в сосудах, становятся заметно меньше.

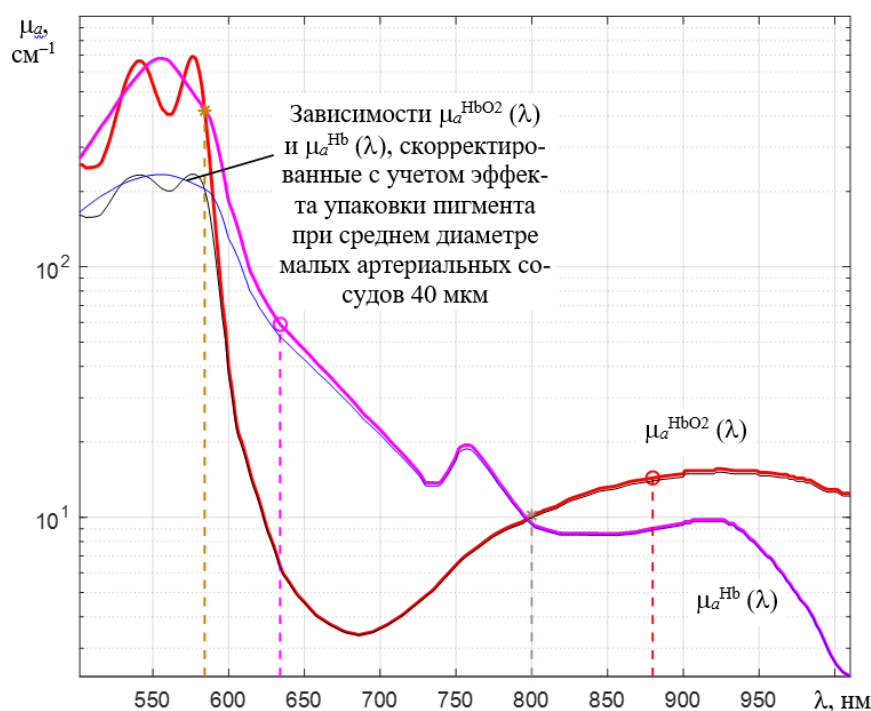


Рис. 1. Натуральные спектральные показатели поглощения излучения HbO_2 и Hb гемоглобином и их скорректированные зависимости при учете фактора упаковки «пигмента» (гемоглобина в эритроцитах и крови в сосудах) при среднем диаметре внутреннего просвета малых пульсирующих артериальных сосудов 40 мкм

Натуральный спектральный показатель поглощения излучения оксигенированного и неоксигенированного гемоглобина рассчитывается через соответствующие табулированные их молярные показатели поглощения $\kappa^{\text{HbO}_2}(\lambda)$ и $\kappa^{\text{Hb}}(\lambda)$, которые можно найти в [4]. Тогда, при обозначении $SaO_2 = S_a$, натуральный показатель поглощения артериальной крови имеет вид

$$\mu^{\text{bla}}(\lambda, C_{mV}^{\text{Hb}}, S_a, C_V^{\text{bla}}) = C_{mV}^{\text{Hb}} C_V^{\text{bla}} \left[S_a \kappa^{\text{HbO}_2}(\lambda) + (1 - S_a) \kappa^{\text{Hb}}(\lambda) \right], \quad (1)$$

где C_V^{bla} – объемная концентрация артериальной крови в единичном объеме ткани (обычно 30 % от концентрации всей крови $C_V^{\text{bl}\Sigma}$), C_{mV}^{Hb} – массовая концентрация общего содержания гемоглобина в крови (около 150 г/л). Зависимости на рис. 2, иллюстрируют влияние сатурации гемоглобина крови на спектральные показатели поглощения крови и ткани.

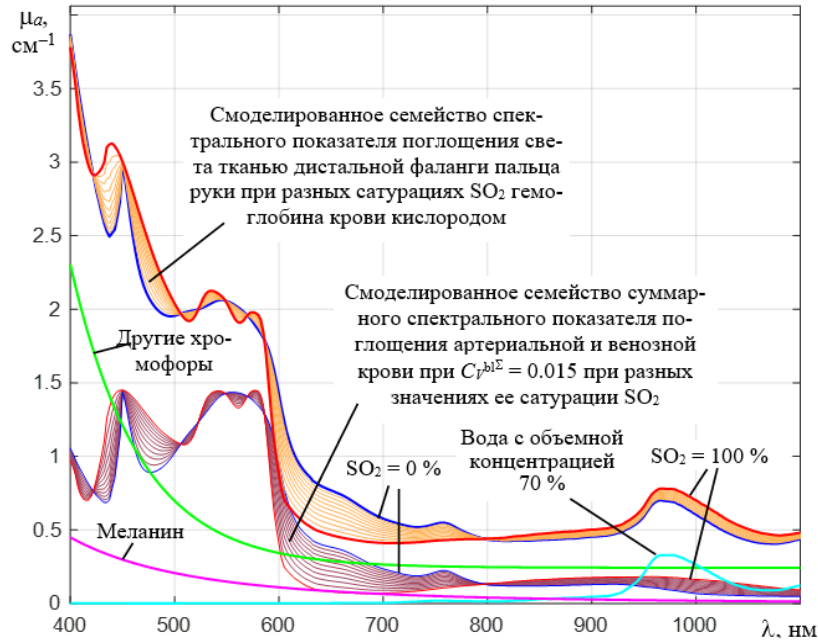


Рис. 2. Приближенные спектральные зависимости показателя поглощения ткани, учитывающие поглощение крови, воды, меланина и других тканей при среднем диаметре сосудов микроциркуляторного русла 29 мкм и изменении сатурации гемоглобина крови кислородом от 0 до 100 % с шагом 10 %

Разделив регистрируемый сигнал на пульсирующую и неппульсирующую составляющие и взяв отношение сигналов, получим выражение

$$\frac{I_{\max}(\lambda_1)}{I_{\min}(\lambda_1)} = 1 + \frac{\Delta I(\lambda_1)}{I_{\min}(\lambda_1)} = \frac{e^{-[S_a \kappa_m^{\text{HbO}_2}(\lambda_1) + (1 - S_a) \kappa_m^{\text{Hb}}(\lambda_1)] C_{mV}^{\text{Hb}} C_V^{\text{bla}}_{\min}}}{e^{-[S_a \kappa_m^{\text{HbO}_2}(\lambda_1) + (1 - S_a) \kappa_m^{\text{Hb}}(\lambda_1)] C_{mV}^{\text{Hb}} C_V^{\text{bla}}_{\max}}}, \quad (2)$$

которое позволяет после дальнейших преобразований получить формулу

для трансмиссионной двухволновой пульсоксиметрии

$$SaO_2 = S_a = \frac{\kappa_m^{Hb}(\lambda_1) - \alpha \kappa_m^{Hb}(\lambda_2)}{\kappa_m^{Hb}(\lambda_1) - \kappa_m^{HbO_2}(\lambda_1) - \alpha [\kappa_m^{Hb}(\lambda_2) - \kappa_m^{HbO_2}(\lambda_2)]}, \quad (3)$$

где $\alpha = \ln[1 + \Delta I(\lambda_1)/I_{\min}(\lambda_1)] / \ln[1 + \Delta I(\lambda_2)/I_{\min}(\lambda_2)]$.

Поступая аналогично для двухволновой рефракционной пульсоксиметрии при описании регистрируемых сигналов с помощью диффузионной функции Грина, можно получить похожее приближенное выражение

$$SaO_2 = S_a = \frac{\kappa_m^{Hb}(\lambda_1) - \alpha k(\lambda_1, \lambda_2) \kappa_m^{Hb}(\lambda_2)}{\kappa_m^{Hb}(\lambda_1) - \kappa_m^{HbO_2}(\lambda_1) - \alpha k(\lambda_1, \lambda_2) [\kappa_m^{Hb}(\lambda_2) - \kappa_m^{HbO_2}(\lambda_2)]}, \quad (4)$$

где $k(\lambda_1, \lambda_2) \approx \sqrt{[\mu'_s(\lambda_2) \mu_a^{bla}(\lambda_1)] / [\mu'_s(\lambda_1) \mu_a^{bla}(\lambda_2)]}$, где $\mu'_s(\lambda)$ и $\mu_a^{bla}(\lambda)$ – спектральные показатели приведенного рассеяния излучения тканью и поглощения излучения артериальной кровью, соответственно.

Спектральную зависимость $\mu'_s(\lambda)$ для пальца руки можно аппроксимировать выражением $\mu'_s(\lambda) = \mu'_s(\lambda_0) \exp[1 - (\lambda/\lambda_0)^{m_\lambda}]$, где $\mu'_s(\lambda_0) \approx 11 \text{ см}^{-1}$, а показатель степени $m_\lambda \approx 1$. Значения $\mu'_s(\lambda)$, как и $\mu_a^{bla}(\lambda)$ в используемом в пульсоксиметрии участке спектра увеличиваются при уменьшении длины волны λ . Это приводит к уменьшению произведения $\alpha k(\lambda_1, \lambda_2)$ и соответствующему увеличению получаемых значений SaO_2 при использовании MAX30102.

Были получены спектрально-временные профили локального коэффициента диффузного отражения $R(\lambda, t, \rho_0)$ ткани дистальной фаланги среднего пальца руки при $\rho_0 = 2,5 \text{ мм}$. Их использование для определения SaO_2 с помощью (3) при $\lambda_2 = 880 \text{ нм}$ и разных значениях λ_1 показало необходимость применения (4) или аппроксимации зависимости $\alpha k(\lambda_1, \lambda_2)$ линейной или квадратичной зависимостью для типичных точек тела.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Методы пульсоксиметрии: возможности и ограничения / Гаранин А. А., Дьячков В. А., Рубаненко А. О. [и др.] // Российский кардиологический журнал. 2023. 28(3S): P. 5467.
2. Accuracy of a New Pulse Oximetry in Detection of Arterial Oxygen Saturation and Heart Rate Measurements: The SOMBRERO Study / Marinari, S., Volpe, P., Simoni, M. [et al.] // Sensors, 2022. P. 22, 5031.
3. Recommended Configurations and Operating Profiles for MAX30101/MAX30102 EV Kits. Maxim Integrated. UG6409; Rev 0; 3/18. P. 36.
4. Tabulated Molar Extinction Coefficient for Hemoglobin in Water, Scott Prahl [Электронный ресурс]. URL: <http://omlc.org/spectra/hemoglobin/summary.html> (дата обращения: 29.03.2024).

IMPLEMENTATION OF THE WATERSHED METHOD FOR SEPARATING NUCLEI OF CANCER CELLS IN FLUORESCENT IMAGES

Y.U. Lisitsa

Belarussian State University, Minsk, Belarus

E-mail: ylisitsa@bsu.by

This article presents the implementation of the watershed method for the segmentation of nuclei in fluorescent images of cancer cells. In this study, we apply the watershed algorithm to fluorescent images of cancer cells, aiming to accurately find the boundaries of nuclei for further analysis. The article discusses the methodology used for preprocessing the images and applying the watershed algorithm. The watershed shows the following drawbacks: watershed lines run through nuclei rather than along their boundaries, merged nuclei persist after processing, over-segmentation leads to noise on the nuclei mask.

Key words: watershed; fluorescent microscopy; image processing.

РЕАЛИЗАЦИЯ МЕТОДА ВОДОРАЗДЕЛА ДЛЯ РАЗДЕЛЕНИЯ ЯДЕР РАКОВЫХ КЛЕТОК НА ФЛЮОРЕСЦЕНТНЫХ ИЗОБРАЖЕНИЯХ

Е.В. Лисица

Белорусский государственный университет, Минск, Беларусь

E-mail: ylisitsa@bsu.by

В данной статье представлена реализация метода водораздела для сегментации ядер на флуоресцентных изображениях раковых клеток. В этом исследовании мы применяем алгоритм водораздела к флуоресцентным изображениям раковых клеток, стремясь точно найти границы ядер для дальнейшего анализа. В статье обсуждается методология предварительной обработки изображений и применения алгоритма водораздела. Метод водораздел имеет следующие недостатки: линии водораздела проходят через ядра, а не вдоль их границ, слитые ядра сохраняются после обработки, чрезмерная сегментация приводит к шуму на маске ядер.

Ключевые слова: водораздел; флуоресцентная микроскопия; обработка изображений.

INTRODUCTION

Fluorescence microscopy plays a pivotal role in cancer research by providing high-resolution imaging of cellular structures. Accurate segmentation of nuclei from fluorescent images is essential for understanding cellular morphology and pathology [1]. One of the most efficient ways of image segmenation is the usage of the artificial neural networks, such as convolutional neural networks. However, as they estimate the probabilities of the nuclei locations in the image, it occurs to the presences of the clumped objects on their

binary masks during the thresholding [2]. The watershed method is a prominent technique for segmenting overlapping objects in digital images. In the context of cancer cell analysis, it offers a promising approach to delineate individual nuclei within densely packed regions [3]. This paper introduces an implementation of the watershed method tailored for separating nuclei of cancer cells in fluorescent images. The aim of the paper is to implement the realization of the watershed segmentation method with markers based on the image preprocessing.

DATA

The article examines three-channel fluorescent images of breast cancer cells, obtained with a 10x magnification on a Nikon TE200 inverted epifluorescence microscope equipped with a Photometrics 300 series CCD camera. The image size is 2048×2048 pixels. The indicator of cancer cells is the estrogen receptor protein labeled with Cy5 dye (registered in the red channel of the image). In the cytoplasm of cancer cells, the cytokeratin protein accumulates, were labeled with Cy3 dye (registered in the green channel). The 4,6-DAPI dye was used to label all nuclei (registered in the blue channel). Accordingly, the markers for cancer cells are two dyes – Cy5 and Cy3[4].

The images were preprocessed using the algorithm described in paper [2] to estimate the probabilities of nuclei locations. The binary masks were obtained by thresholding the probability images, with a threshold of 0.5.

WATERSHED ALGORITHM

Let I be the input image, where $I(x,y)$ represents the intensity at pixel coordinates (x,y) . We consider I as a topographic surface.

1. **Gradient Calculation:** Compute the gradient magnitude of the image to identify potential watershed lines. This can be represented mathematically as:

$$|\nabla I| = \sqrt{(\partial I / \partial x)^2 + (\partial I / \partial y)^2}$$

2. **Marker Generation:** Identify markers for the watershed transformation. These markers typically correspond to the local minima of the gradient magnitude image. Let M represent the marker image, where $M(x,y)=1$, if pixel (x,y) is a marker, and $M(x,y)=0$ otherwise.

3. **Flood-Filling:** Simulate a flooding process starting from the markers. This process is akin to placing water at each marker and letting it fill the image. At each step, the water from neighboring pixels flows towards the marker with the lowest intensity. This process continues until all pixels are assigned to a marker. Mathematically, this can be represented using a flooding function F defined recursively as: $F(x,y) = (x',y') \in N(x,y) \min\{F(x',y') + I\}$, where $N(x,y)$ is

the set of neighboring pixels of (x,y) . The initial values of F are set to 0 at marker locations.

4. Watershed Line Formation: Watershed lines form where the flooding fronts from different markers meet. These lines delineate the boundaries between segmented regions. They can be represented as a binary image W , where $W(x,y)=1$ indicates a watershed line pixel and $W(x,y)=0$ otherwise.

Watershed segmentation involves computing the gradient of the image, identifying markers, simulating a flooding process from these markers, and finally extracting watershed lines to segment the image into regions [5].

IMPLEMENTATION

The flow-chart of image processing for watershed implementation is shown in the figure 1.

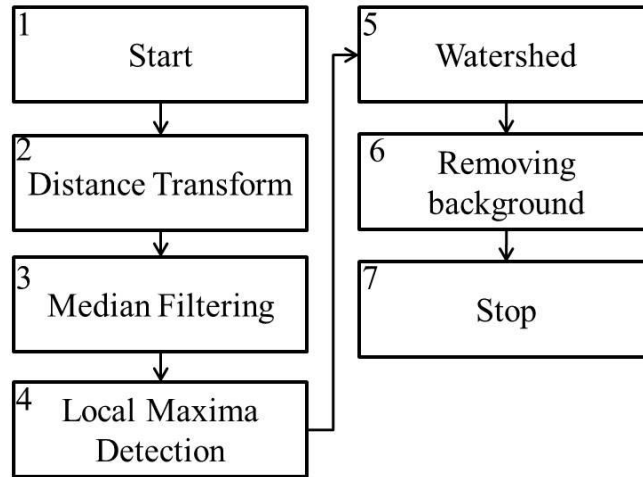


Fig. 1. Watershed implementation flow-chart

1. **Start.** The input consists of a mask of clumped nuclei ($nuclei_{clumped}$), which represents an input image containing clumped nuclei, and the minimum area of nuclei ($area_{min}$).

2. **Distance Transform.** The code computes the distance transform of $nuclei_{clumped}$ using Euclidean distance transform, storing the result in $image_{transform}$.

3. **Median Filtering.** It calculates the radius (r_{min}) based on the $area_{min}$. Then, it applies median filtering to $image_{transform}$ using a square-shaped structuring element with a size determined by r_{min} .

4. **Local Maxima Detection.** It identifies the coordinates of peak local maxima in the negated $image_{transform}$. Then it uses a 3x3 neighborhood and assigns labels based on $nuclei_{clumped}$.

5. **Watershed.** It initializes a mask based on the peak local maxima coordinates and performs watershed segmentation on the negated distance

transform image using these markers. The resulting segmented image is stored in *labels*.

6. **Removing background.** Finally, it removes the background (regions labeled as 0 in *labels*) from *nuclei_{clumped}*, isolating the individual nuclei.

The implementation is made on the Python language and it includes following libraries: numpy, skimage, scipy.

RESULTS

The results of segmentation are shown in the figure 2. The grayscale image (Fig. 2A) is presenting the estimations for the nuclei probability locations. The figure 2B shows only clumped nuclei after thresholding. The distance transform image after filtration is presented in the figure 2C. The separated nuclei by the watershed method are shown in the image 2D.

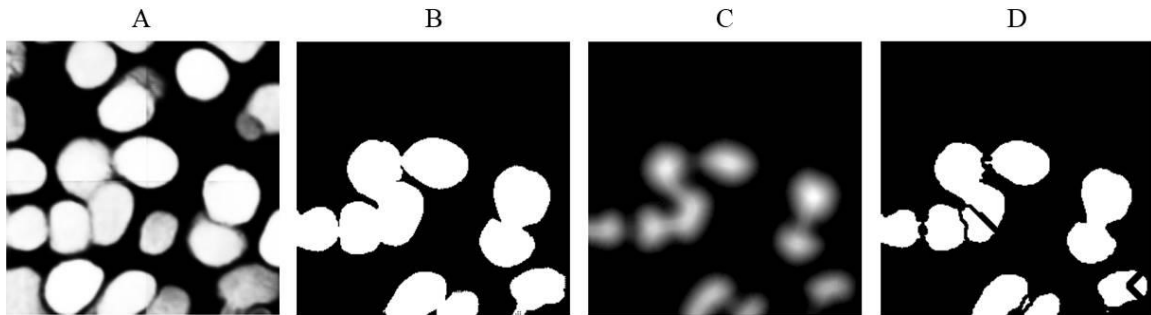


Fig. 2. Segmentation results:

A - Grayscale image; B - Clumped nuclei; C - Distance transform image after filtration;
D - Separated nuclei

CONCLUSION

As indicated by the segmentation results, the watershed method allows for separating merged objects. Even though our implementation leverages image processing techniques to enhance segmentation accuracy and efficiency, the disadvantages of the method include:

1. Watershed lines do not run along the boundaries of nuclei but through the nuclei themselves.
2. Merged nuclei remain after processing with the method.
3. Over segmentation, this also results in noise presence on the nuclei mask.

REFERENCES

1. Molecular Biology of the Cell / B. Alberts [et al.] // New York: Garland Science, 2012. P. 1392.
2. *Silun X., Skakun, V.* Fully-automated segmentation of tumor nuclei in cancer tissue images // Pattern Recognition and Information Processing (PRIP'2021) : Proceedings of the 15th International Conference, 21-24 Sept. 2021, Minsk, Belarus. Minsk : UIIP NASB, 2021. P. 168-172.
3. *R. C. Gonzalez, R. E. Woods, S. L. Eddins.* Gonzalez, R. C. Digital Image Processing Using MATLAB // Gatesmark Publishing, 2020.
4. *Е. В. Лисица, Н. Н. Яцков, В. В. Скакун, П. Д. Кривошеев, В. В. Апанасович.* Отбор информативных геометрических признаков ядер клеток на люминесцентных изображениях раковых клеток // Информатика. 2019. Т. 16, № 2. С.7-17.
5. *L. Najman, M. Schmitt* Watershed of a continuous function // Signal Processing (Special issue on Mathematical Morphology. 1994. Vol. 38, P. 99–112.

ИССЛЕДОВАНИЕ МЕТОДОВ СНИЖЕНИЯ РАЗМЕРНОСТИ В ХОДЕ АНАЛИЗА ДАННЫХ ФЛУОРЕСЦЕНТНОЙ СПЕКТРОСКОПИИ

А.В. Ромашко, Н. Н. Яцков

Белорусский государственный университет, Минск, Беларусь
E-mail: alexey.romashko2002@gmail.com

В работе представлены результаты исследования методов снижения размерности, применяемых к анализу данных флуоресцентной спектроскопии. Выполнен анализ алгоритмов KernelPCA, t-SNE, UMAP, автокодировщик и ICA на имитационно смоделированных наборах кривых затухания флуоресценции молекулярных систем различной сложности. Наивысшие скорость вычислений и точность восстановления кластеров данных демонстрируют методы t-SNE, UMAP и KernelPCA.

Ключевые слова: флуоресцентная спектроскопия; затухание флуоресценции; имитационное моделирование; методы снижения размерности.

ВВЕДЕНИЕ

Методы экспериментальной флуоресцентной спектроскопии интенсивно применяются в физико-химических, биофизических и биомедицинских исследованиях [1]. Постоянное развитие методов обусловлено разработкой эффективных молекулярных флуорофоров, в том числе генетически экспрессированных белков (например, GFP), оптических систем лазерного возбуждения и регистрации излучения, позволяющих проводить высокоточные измерения, развитием компьютерных технологий хранения и обработки данных. Новые экспериментальные высокопроизводительные методы регистрации времени затухания флуоресценции позволяют получать большие данные, характеризующиеся высоким спектральным, временным и пространственным разрешением [2]. Результативность анализа больших наборов кривых затухания флуоресценции может быть повышена при применении эффективных методов снижения размерности данных [3].

Целью работы является сравнительный анализ наиболее эффективных методов снижения размерности на смоделированных данных флуоресцентной спектроскопии, кривых затухания флуоресценции.

МЕТОДЫ МОДЕЛИРОВАНИЯ И АНАЛИЗА ДАННЫХ

Имитационное моделирование кривых затухания флуоресценции

Для всестороннего исследования алгоритмов снижения размерности данных необходимо смоделировать кривые затухания флуоресценции мо-

лекулярных систем различной сложности [4]. Рассмотрим одноэкспоненциальную, трёхэкспоненциальную, стрэтч, или «растянутую», модели кривых затухания флуоресценции. Имитационно смоделированы 3 набора кривых затуханий флуоресценции с параметрами: 1) $N = 300$ (число кривых затуханий флуоресценции), $K = 1200$ (число временных каналов регистрации флуоресценции); 2) $N = 1200$, $K = 1200$ и 3) $N = 1500$, $K = 1200$. Параметр, характеризующий уровень шума, $\sigma = 0,2$, временной интервал 10^{-8} с.

Методы снижения размерности данных

Для практического исследования выбраны наиболее популярные и эффективные методы снижения размерности данных, такие как KernelPCA, ICA, UMAP, t-SNE и нейросеть автокодировщик [3, 5, 6].

Метод главных компонент ядра (KernelPCA) – метод понижения размерности данных, использует функции ядра для работы с нелинейными зависимостями в данных.

Анализ независимых компонент (ICA) – метод, определяющий независимые компоненты в данных, разделяя исходные сигналы на несмешанные и некоррелированные составляющие.

Стохастическое вложение соседей с t-распределением (t-SNE) – нелинейный метод, преобразует расстояния между точками в вероятности соседства и затем оптимизирует местоположение точек в новом пространстве, чтобы максимизировать подобие соседства.

Uniform manifold approximation (UMAP) – нелинейный метод, в основе которого лежат преобразования расстояний между наблюдениями в граф соседства и минимизация различий между расстояниями в исходном и преобразованном пространствах.

Автокодировщик – это нейронная сеть автоэнкодера, которая используется для понижения размерности данных путем обучения слоев кодировщика и декодировщика сети автоэнкодера.

Вычислительные алгоритмы реализованы на языке программирования Python с использованием библиотек NumPy, Pandas, Matplotlib и Tensorflow.

Критерии оценивания эффективности методов снижения размерности

Для оценки качества алгоритмов снижения размерности данных используется идея группировки или кластеризации наборов смоделированных данных с известными параметрами [7-9]. В этом случае основной характеристикой эффективности методов снижения размерности данных может быть точность восстановления кластеров смоделированных данных. Для оценки качества кластеризации взяты меры точности *Accuracy*

(далее Асс, в долях от 1) отнесения интенсивностей флуоресценции к соответственным кластерам и индекса разделения кластеров d [10].

В качестве оценки производительности алгоритмов рассмотрено время выполнения вычислений t .

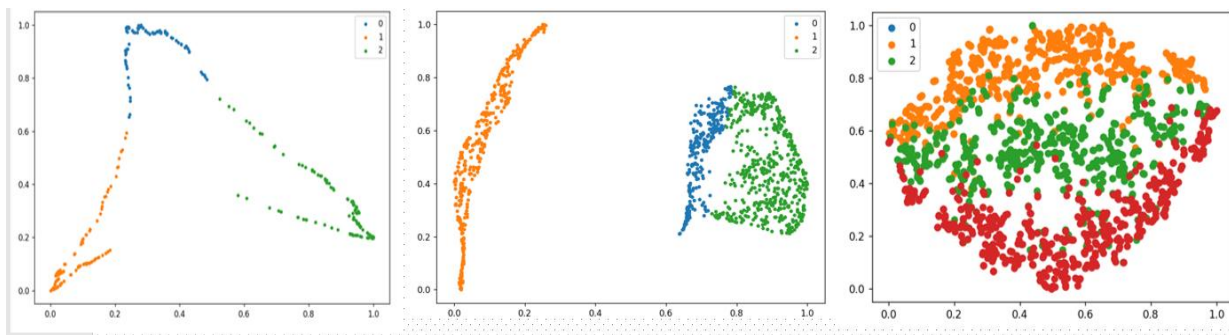
РЕЗУЛЬТАТЫ

Выполнен сравнительный анализ методов снижения размерности на трех наборах смоделированных данных. Результаты сравнительного анализа представлены на рисунке и в таблице.

Оценки эффективности алгоритмов снижения размерности данных

	ICA	t-SNE	UMAP	Автокод.	KernelPCA
Система 1					
t , с	0,13*	0,59	1,97	2,64	0,04
d	0,37	0,52	0,29	0,24	0,27
Асс	0,96	0,97	0,96	0,96	0,95
Система 2					
t , с	0,10	4,23	6,56	3,35	0,25
d	0,46	0,72	0,55	0,47	0,55
Асс	0,73	0,86	0,86	0,79	0,83
Система 3					
t , с	0,10	3,41	8,08	2,56	0,53
d	1,73	1,37	1,39	2,59	1,42
Асс	0,75	0,96	0,90	0,85	0,95

Примечания. Среднеквадратические отклонения в оценках точности не превышали 5% значений от их абсолютных величин



Оцененные кластеры смоделированных данных в пространствах двух наиболее информативных координат. Слева-направо – система 1, KernelPCA; система 2, UMAP; Система 1, t-SNE

В ходе сравнительного анализа установлено, что ICA является самым быстрым методом. KernelPCA показал в целом более длительное время, но гораздо точнее справился на сложных системах, чем сравниваемые алгоритмы. UMAP и t-SNE демонстрируют высокую точность и хорошую

визуализацию, однако время затратные. Автокодировщик сильно зависит от тренировочной выборки, показывает неплохие результаты, но уступает t-SNE и UMAP.

ЗАКЛЮЧЕНИЕ

Выполнено исследование методов снижения размерности данных KernelPCA, UMAP, t-SNE, автокодировщика и ICA для трех смоделированных молекулярных систем, представленных кривыми затухания флуоресценции. Наилучшие результаты получены для методов KernelPCA, t-SNE и UMAP – точность восстановления наборов данных смоделированных кластеров – 0,97, 0,86 и 0,96. KernelPCA хуже разделяет кластеры в более сложных системах, однако более быстрый – время 0,04, 0,25 и 0,53 с для трех систем соответственно. UMAP и t-SNE лучше разделяют кластеры, эффективны при визуализации данных и анализе сложных систем.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. *Лакович Дж.* Основы флуоресцентной спектроскопии. Пер. с англ. // М.: Мир, 1986. С. 496.
2. *Yatskou M. M., Apanasovich V. V.* Simulation modelling and data mining approach for the study of applied fluorescence spectroscopy systems // *Journal of the Belarusian State University. Physics.* 2024. Vol. 1. P. 4–15.
3. *Hinton G. E., Salakhutdinov R. R.* Reducing the dimensionality of data with neural networks // *Science.* 2006. Vol. 313, N. 5786. P. 504–507.
4. *Яцков Н.Н., Скакун В.В., Апанасович В. В.* Метод обработки кинетических кривых затухания флуоресценции с использованием алгоритмов интеллектуального анализа данных // *Журнал прикладной спектроскопии.* 2020. Т.87, №2. С. 320-333. DOI 10.1007/s10812-020-01004-3
5. *McInnes L., Healy J.* UMAP: Uniform manifold approximation and projection for dimension reduction // *ArXiv.* 2020. arXiv:1802.03426. P. 1-63. <https://doi.org/10.48550/arXiv.1802.03426>
6. *van der Maaten L.J.P., Hinton G.E.* Visualizing high-dimensional data using t-SNE // *Journal of Machine Learning Research.* 2008. Vol. 9. P. 2579-2605.
7. Complex analysis of fluorescence intensity fluctuations of molecular compounds / *Yatskou M.M., Skakun V.V., Nederveen-Schippers L. [et al.]* // *Journal of Applied Spectroscopy.* 2020. N87(4). P. 685–692. DOI: 10.1007/s10812-020-01055-6
8. *Яцков Н. Н., Апанасович В. В.* Комплексный анализ данных при исследовании сложных биомолекулярных систем // *Информатика.* 2021. Т. 18, № 1. С. 105–122. DOI: <https://doi.org/10.37661/1816-0301-2021-18-1-105-122>
9. Component analysis of photon counting histograms in fluorescence fluctuation spectroscopy experiments / *Skakun V.V., Yatskou M.M., Nederveen-Schippers L. [et al.]* // *Journal of Applied Spectroscopy.* 2022. N89(5). P. 930–939. DOI: 10.1007/s10812-022-01450-1
10. *Сивоголовко Е.В.* Методы оценки качества чёткой кластеризации // *Компьютерные инструменты в образовании.* 2011. №4. С. 15-31.

МНОГОКАНАЛЬНЫЙ ИНТЕЛЛЕКТУАЛЬНЫЙ ИММУНОФЕРМЕНТНЫЙ АНАЛИЗ НА ОСНОВЕ ОПТИЧЕСКИХ МИКРОРЕЗОНАТОРОВ

А.В. Саечников, Э.А. Чернявская, В.А. Саечников

*Белорусский Государственный Университет, Минск, Беларусь
E-mail: anton.saetchnikov@gmail.com*

В данной работе предлагается биочип на основе оптического микрорезонатора, предназначенный для иммуноферментного анализа в сложных растворах. Он использует метод мультиплексного параллельного опроса микрорезонаторов и вычислительные методы на основе машинного обучения для уменьшения зависимости от постоянства производства сенсорных единиц.

Ключевые слова: оптический микрорезонатор; машинное обучение; детектирование; мода шепчущей галереи.

Оптические биосенсоры основаны на взаимодействии между оптическим полем и биохимическими веществами, обеспечивая компактность, доступность, высокую чувствительность и мультиплексное обнаружение для комплексного анализа биохимических систем. Датчики, использующие оптические микрорезонаторы, обеспечивают взаимодействие оптической волны с образцами. Оптическое поле локализуется в кольцевых полостях, что позволяет волнам распространяться по периферии резонатора и взаимодействовать друг с другом для создания мод шепчущей галереи (МШГ) [1]. Известные высоким показателем добротности, они обеспечивают чувствительность до уровня единичных молекул [2]. Множество микрорезонаторов может быть плотно упаковано в небольших областях или объемах для создания массива [3,4], где применение машинного обучения позволяет проводить измерения с использованием доступного оборудования и производить количественные оценки больших данных [5,6].

Критическим требованием для сенсора, поддерживающим машинное обучение, является необходимость высокой согласованности в производстве, где вариабельность чипов может существенно снизить производительность из-за тонких изменений в их параметрах.

В этой статье мы сообщаем об интеллектуальном биочипе на основе оптических микрорезонаторов обеспечивающем иммуноферментный анализ в сложных биохимических растворах. Он объединяет технологию мультиплексного микрорезонаторного детектирования вместе с методами машинного обучения, что позволяет снизить требования к стабильности изготовления чипов и тем самым значительно расширить реальный потенциал биодетектирования на основе МШГ.

МАТЕРИАЛЫ И МЕТОДЫ

Чип объединяет тысячи стеклянных микросфер, зафиксированных на тонком покровном стекле, образуя отдельные измерительные каналы. Он может содержать до 16 каналов, с возможностью расширения до 100 каналов за счет использования микрорезонаторов малого размера. Оптимизация толщины клеевого слоя снижает потери на связь и улучшает чувствительность отдельных датчиков, максимизируя коэффициент добротности под нагрузкой, приближающийся к 10^6 в воде.

В качестве источника излучения использовался одночастотный лазер с длиной волны 785 нм. Регистрируемый сигнал представляет собой время зависимый набор интенсивностей, которые изменяются в соответствии с спектральными характеристиками каждого резонатора, такими как спектральное положение и ширина лазерной линии.

РЕЗУЛЬТАТЫ

Простота изготовления и метода опроса имеет недостаток: результаты измерений биочипов различаются по размерности и механизмам, лежащим в основе отклика. Для обеспечения согласованности данных различных чипов необходимо построить вектор признаков из измеренных данных, чтобы повысить устойчивость зондирования. Помимо нормализации и согласования масштаба, полученные данные обрабатываются с помощью анализа главных компонент, который является общим методом уменьшения размерности, преобразуя исходные переменные в независимые компоненты, известные как главные компоненты (ГК).

Несмотря на нелинейную и уникальную функции сигнала для каждой микросферы можно точно получить образ спектрального сдвига в первой главной компоненте. Это становится возможным при наличии достаточного количества (> 100) одновременно доступных откликов резонаторов. В результате набор первых ГК для функциональных каналов используются в качестве исходных признаков для обучения нейронной сети.

В рамках работы было изготовлено несколько биочипов, каждый из которых использовался несколько раз для демонстрации определения типа и концентрации иммуноглобулинов. Чипы содержали по 4 функциональных канала, где два использовались как контрольные, а остальные два обрабатывались антителами anti-human IgG и anti-rabbit IgG.

Избирательное обнаружение различных составляющих сложных растворов с концентрациями отдельных иммуноглобулинов от 100 до 50000 нг/мл достигалось путем связывания с антителами на поверхности микрорезонаторов. Отклик на среду с растворенными иммуноглобулинами измеренный в течение 2000 с совместно с реакцией на буферный раствор до

и после измерения аналита использовался как исходный набор данных для обучения. После второго измерения буферного раствора чип подвергался процедуре регенерации. Набор данных, построенный таким образом, позволяет учесть изменения в эффективности связывания, вариации чувствительности среди отдельных образцов многоканальных сенсоров, а также воспроизводимость процесса осаждения функционального слоя.

Было показано, что простая однопроходная нейросеть с сверточными и полносвязными слоями подходит для решения задачи предсказания состава растворов и концентрации отдельных компонент. Сеть с оптимизированной архитектурой показала высокую точность ($> 98\%$) для определения типа иммуноглобулина с относительной ошибкой ~ 10 нг/мл.

ЗАКЛЮЧЕНИЕ

Мы разработали и охарактеризовали биочип на основе оптического микрорезонатора, предназначенный для иммуноферментного анализа в сложных средах. Этот метод основывается на преимуществах методики одновременной регистрации сигнала для тысяч микрорезонаторов, а также методах машинного обучения.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. M. Loyez, M. Adolphson, J. Liao, L. Yang. From Whispering Gallery Mode Resonators to Biochemical Sensors // ACS sensors. 2023. 8. P. 2440.
2. Whispering-Gallery Mode Optoplasmonic Microcavities. / M. C. Houghton, S. V. Kashanian, T. L. Derrien [et al.] // ACS Photonics. 2024. 86. P. 37.
3. Saetchnikov A. V., Tcherniavskaia E. A., Saetchnikov V. A., Ostendorf, A. Multiresonator imaging sensor for the aerial parameters detection // IEEE Journal on Miniaturization for Air and Space Systems. 2021. 2(2). P. 84–91.
4. Saetchnikov A. V., Tcherniavskaia E. A., Saetchnikov V. A. Ostendorf, Detection of per- and polyfluoroalkyl water contaminants with multiplexed 4D microcavities sensor // Photonics Res. 2023. 11. P. A88.
5. Saetchnikov A. V., Tcherniavskaia E. A., Saetchnikov V. A., Ostendorf, A. Deep-learning powered whispering gallery mode sensor based on multiplexed imaging at fixed frequency // Opto-Electronic Adv. 2020. 3. P. 200048.
6. Saetchnikov A. V., Tcherniavskaia E. A., Saetchnikov V. A., Ostendorf, A. Intelligent optical microresonator imaging sensor for early stage classification of dynamical variations // Advanced Photonics Research. 2021. 7. P. 2100242.

ОПРЕДЕЛЕНИЕ ОКСИГИНАЦИИ КРОВИ КОРЫ ГОЛОВНОГО МОЗГА

А.В. Лебедевский, В.А. Фираго, М.М. Кугейко, Н.В. Левкович,
С. Г. Славинский, К.И. Шулико

Белорусский Государственный Университет, Минск, Беларусь
E-mail: firago@bsu.by

Приведены схемы устройств для неинвазивного определения оксигенации крови коры головного мозга. Для измерения сатурации артериальной крови предлагается микропроцессорное устройство на модуле пульсоксиметра МАХ30102. Показано, что для снижения влияния индивидуальных особенностей тканей головы на результаты определения тканевой сатурации StO_2 требуется получение спектрально-пространственных профилей диффузного отражения с помощью подвижного микроспектрометра С11708МА.

Ключевые слова: кора головного мозга; рефракционная пульсоксиметрия; сатурация артериальной крови; тканевая сатурация; спектры диффузного отражения

ВВЕДЕНИЕ

Неинвазивное определение оксигенации при использовании отношений количества молекул оксигенированного m^{HbO_2} и неоксигенированного m^{Hb} гемоглобина в артериальной $SaO_2 = S_a = m_{Va}^{HbO_2} / (m_{Va}^{HbO_2} + m_{Va}^{Hb})$ и венозной $SvO_2 = S_v = m_{Vv}^{HbO_2} / (m_{Vv}^{HbO_2} + m_{Vv}^{Hb})$ крови коры головного мозга, а также тканевой сатурации $StO_2 = S_t = (C_V^{bla} S_a + C_V^{blv} S_v) / C_V^{bl\Sigma}$, где C_V^{bl} – соответствующие объемные концентрации, возможно по спектрам диффузного отражения светового излучения тканями головы [1]. Если у детей влияние костной ткани черепа на измеряемые значения StO_2 небольшое, для взрослых необходимы более сложные методы измерений [2].

Применение разработанной нами методики регистрации спектрально-пространственных профилей локального коэффициента диффузного отражения поверхностных биотканей в широком участке спектра от 490 до 1150 нм и методики их обработки на основе использования диффузионного приближения [3, 4] должно упрощать определение оптических характеристик тканей головы и последующее вычисление значения StO_2 . Выяснено, что при неопределенности спектральных показателей поглощения основных хромофоров тканей головы (из-за влияния индивидуальных особенностей), следует для повышения устойчивости получаемых результатов решения обратной задачи проводить параллельные измерения SaO_2 двухволновым пульсоксиметрическим способом. Поэтому в докладе рассматриваются вопросы

определения SaO_2 артериальной крови коры головного мозга и регистрации спектров диффузного отражения тканей головы.

ТЕХНИКА ДЛЯ ОПРЕДЕЛЕНИЯ САТУРАЦИИ КРОВИ КОРЫ МОЗГА

Как видно из схемы блока определения SaO_2 артериальной крови коры головного мозга, представленной на рис. 1, в нем используется модуль пульсоксиметра MAX30102, который дополнен двумя парами внешних сверхъярких светодиодов (R с максимумом спектра излучения $\lambda = 660$ нм и IR с максимумом при $\lambda = 880$ нм) и соответствующей схемой их управления. Управление работой этого блока осуществляет микропроцессор STM32F103C8T6, который по USB интерфейсу связан с компьютером, отображающим получаемые зависимости на экране монитора. Линии SCL, SDA используются для обмена данными по I²C шине между MAX30102 и STM32, а линия ALL_LED_ON служит для включения и выключения внешних светодиодов LED1–LED4.

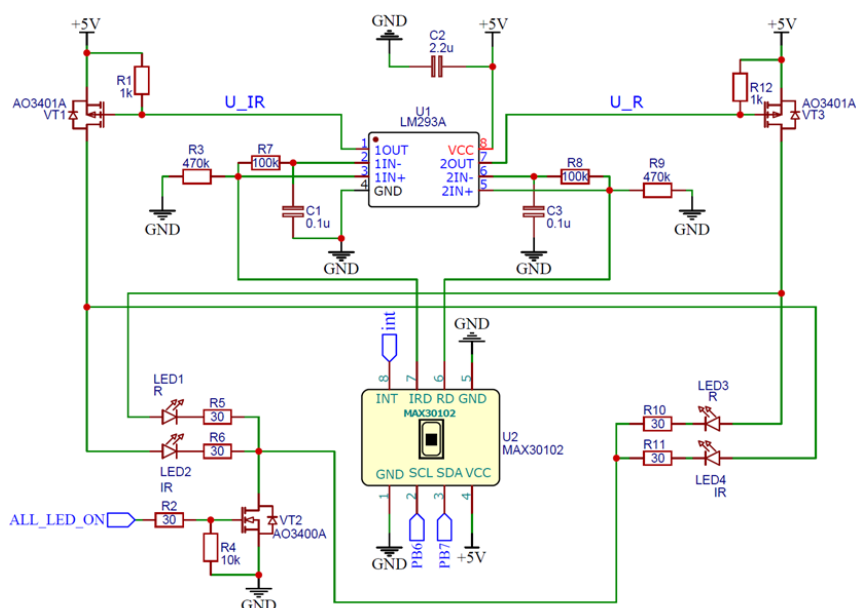


Рис. 1. Принципиальная электрическая схема блока измерения SaO_2 артериальной крови головного мозга на основе модуля MAX30102

Модуль MAX30102 имеет 2 вывода RD и IRD драйвера управления токами встроенных светодиодов R и IR, на которых при их включении появляются соответствующие импульсы с регулируемой длительностью в диапазоне 69...411 мкс. С помощью двух компараторов они преобразуются в логические импульсы U_R и U_IR, которые поочередно включают внешние R и IR светодиоды. Такое решение позволяет путем вычитания из сигналов, регистрируемых при одновременной подсветке ткани встроенными и внешними светодиодами, сигналов, которые регистрируются

при подсветке только встроенными светодиодами, разделить между собой пульсоксиметрические сигналы, получаемые при рассеянии излучения светодиодов кожей головы и корой головного мозга. Вклад костной ткани черепной коробки в амплитуду пульсаций потоков, регистрируемых от коры головного мозга, отсутствует, что позволяет, используя двухволновую пульсоксиметрию, определять значение SaO_2 .

Измерения спектральных коэффициентов локального отражения поверхностных тканей $R(\lambda)$, проведенные с помощью оптоволоконного зонда FCR-7UVIR400-2-BX/ME и спектрометра AvaSpec 2048WL и представленные на рис. 2, указывают на уменьшение вклада крови кожи лба в общий спектр поглощения, вследствие отсутствия крови в костной ткани. Расстояние между центром торца центрального световода, подводящего излучение галогенной лампы к ткани, и центрами 6 приемных световодов примерно 650 мкм, поэтому практически фиксировалось излучение, рассеиваемое только кожной и костной тканью лба.

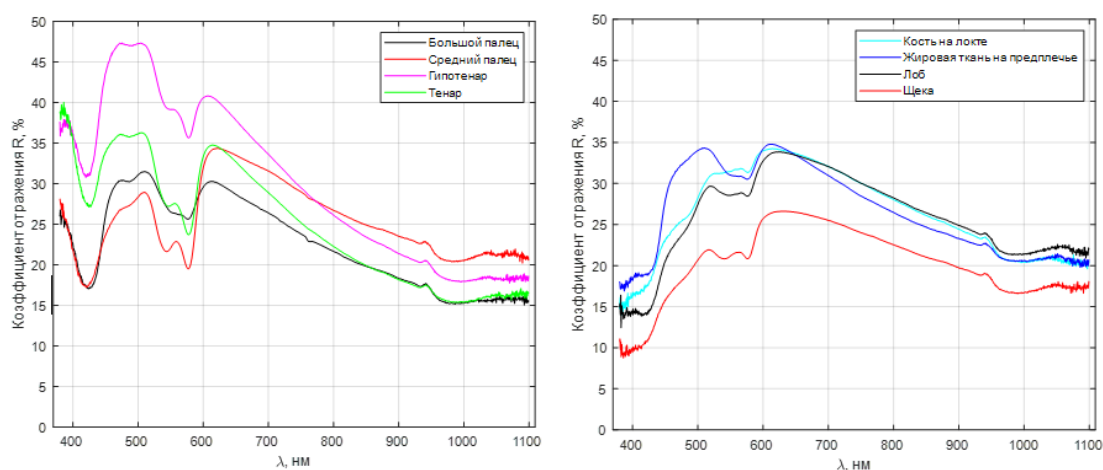


Рис. 2. Зарегистрированные спектральные коэффициенты локального диффузного отражения тканей руки и головы

Расстояние между встроенными в МАХ30102 светодиодами и фотодиодом 4 мм. Поэтому глубина проникновения излучения небольшая, что позволяет определять SaO_2 артериальной крови только кожной ткани. Использование сигналов от внешних светодиодов, которые располагаются на расстоянии 2 см от фотодиода МАХ30102 дает возможность определять SaO_2 артериальной крови коры головного мозга. Действительно, вклад поглощения крови, находящихся в сосудах кожи, небольшой по сравнению с поглощением крови коры головного мозга, тем более при предлагаемой схеме измерений его можно учесть.

Для определения тканевой сатурации StO_2 требуется получение спектрально-пространственных профилей $R(\lambda_i, \rho_k)$ локального коэффициента

диффузного отражения при расстояниях между входной щелью спектрометра и точками ввода излучения не менее 2 см. Предлагаемая схема блока приведена на рис. 3. Для регистрации $R(\lambda_i, \rho_k)$ используется микроспектрометр C11708MA, который имеет объем всего несколько см³ и регистрирует спектры в диапазоне от 580 до 1100 нм. Излучение галогенных ламп подводится оптоволоконными, а изменение расстояния ρ_k между излучателями и центром входной щели C11708MA осуществляется с помощью перемещения корпуса C11708MA миниатюрным шаговым двигателем вдоль кожи головы на расстояние ± 5 мм от центрального положения.

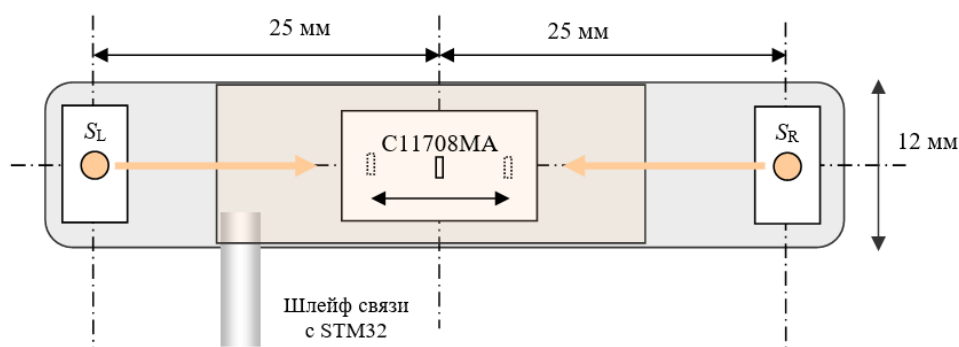


Рис. 3. Схема размещения элементов блока измерения профиля $R(\lambda_i, \rho_k)$ тканей головы на основе подвижного микроспектрометра C11708MA Hamamatsu

Таким образом, совокупность предлагаемых технических решений и методики проведения измерений и их обработки [3, 4] позволит снизить влияние индивидуальных особенностей тканей головы на результаты неинвазивного определения сатураций SaO_2 и StO_2 коры головного мозга.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Noninvasive measurement of cerebral hemoglobin oxygen saturation using two near infrared spectroscopy approaches / Quaresima V., Sacco S., Totaro R. [et al.] // Journal of Biomedical Optics 2000. Vol. 5, № 2. P. 201–205.
2. Absolute quantification of cerebral tissue oxygen saturation with multidistance broadband NIRS in newborn brain / Kovacsova Z., Bale G., Mitra S. [et al.] // Biomedical Optics Express. 2021. Vol. 12, № 2. P. 907–925. 1.
3. Possibilities of Diffuse Reflectance Spectroscopy in Determining and Operational Control of the Optical Properties of Finely Dispersed Scattering Media / Hotra O., Firago V., Shuliko K. [et al.] // Electronics. 2023. Vol. 12, № 13, 2893.
4. *Фираго В.А.* Оценка жесткости малых артериальных сосудов поверхностных биотканей по их спектрально-временным профилям диффузного отражения светового излучения // Журнал прикладной спектроскопии. 2024. Т. 91, № 1. С. 107–123.

НЕИНВАЗИВНАЯ ОЦЕНКА СОСТОЯНИЯ АРТЕРИАЛЬНЫХ СОСУДОВ МИКРОЦИРКУЛЯТОРНОГО РУСЛА

**В.А. Фираго, А.В. Лебедевский, Н.В. Левкович,
С. Г. Славинский, К.И. Шулико, Н.С. Гвоздев, Г.В. Панченко**

Белорусский Государственный Университет, Минск, Беларусь

E-mail: firago@bsu.by

Рассмотрена проблема оценки состояния артериальных сосудов микроциркуляторного русла, которые первыми ощущают негативное воздействие механизмов «раннего старения» системы кровообращения. Приводится состав компьютеризированного комплекса, предназначенного для оперативного неинвазивного обследования состояния артериальных сосудов при ежегодной диспансеризации взрослого населения.

Ключевые слова: тонус артериальных сосудов; скорость распространения пульсовой волны; фотоплетизмограмма; спектры диффузного отражения биотканей

ВВЕДЕНИЕ

В 2022 году вышло Постановление Минздрава РБ о проведении ежегодной диспансеризации взрослого населения старше 39 лет. Его главная цель – определение факторов риска развития заболеваний на ранней стадии. Для эффективного осуществления диспансеризации необходимо использовать комплекс аппаратуры, позволяющий оперативно (в поточном режиме) проводить оценку основных диагностических параметров организма человека.

Сердечно-сосудистые заболевания лидируют по количеству смертей в мире. При своевременной диагностике этих заболеваний можно существенно снижать смертность и затраты на их лечение. Специализированной аппаратуры для проведения оценки состояния сердечнососудистой системы, кроме обычно принятых исследований (частоты сердечных сокращений, измерения систолического и диастолического давления крови, ЭКГ, флюорографии, общего анализа крови и ее биохимического анализа), на текущий момент в клиниках РБ нет. Поэтому одним из актуальных направлений при создании такого комплекса аппаратуры может быть разработка компьютеризированного устройства для оперативной неинвазивной оценки состояния артериальных сосудов.

АРТЕРИАЛЬНОЕ ДАВЛЕНИЕ И ЖЕСТКОСТЬ АРТЕРИАЛЬНЫХ СОСУДОВ

Эффективная работа сердечно-сосудистой системы обеспечивается не только сократительной способностью миокарда сердца и транспортной функцией артерий, но и благодаря демпфирующей функции аорты и магистральных артерий. Они, вследствие своей высокой эластичности, способны,

растягиваясь, принимать выбрасываемый сердцем объем крови, перенося таким образом значительную часть энергии сердечного сокращения на период диастолы, когда кровь постепенно выдавливается из них в периферические сосуды. Это приводит к снижению систолического артериального давления и повышению диастолического давления, тем самым уменьшая повреждающее действие пульсовой волны на сосуды мозга, сердца и почек, хотя кровоснабжение их при этом улучшается. Поэтому значение жесткости стенок магистральных артерий, оцениваемая по скорости распространения пульсовой волны (СРПВ), согласно Европейским рекомендациям по артериальной гипертензии 2013 года является признаком кардиоваскулярного риска. При превышении СРПВ значения 12 м/с этот риск существенно повышается, поскольку обычно СРПВ не превышает 5–6 м/с.

Высокая распространенность сердечно-сосудистых заболеваний требует выявления предрасположенности к ним пациентов среднего и даже молодого возраста. Причем, при диспансеризации необходимо особое внимание уделять оценке состояния малых артериальных сосудов, которые первыми ощущают негативное воздействие развивающихся механизмов «раннего старения» сосудов. При этом содержание эластина в стенках сосудов падает при одновременном увеличении концентрации коллагена, что приводит к увеличению жесткости стенок артериальных сосудов и, как следствие, к повышению артериального давления и скорости распространения пульсовой волны как по магистральным, так и периферическим артериальным сосудам. При уточняющей диагностике гипертензии выделяют: стойкое сужение артериол (систолическое и диастолическое давления оба увеличены) и атеросклероз крупных артериальных сосудов (систолическое увеличенное, диастолическое в норме).

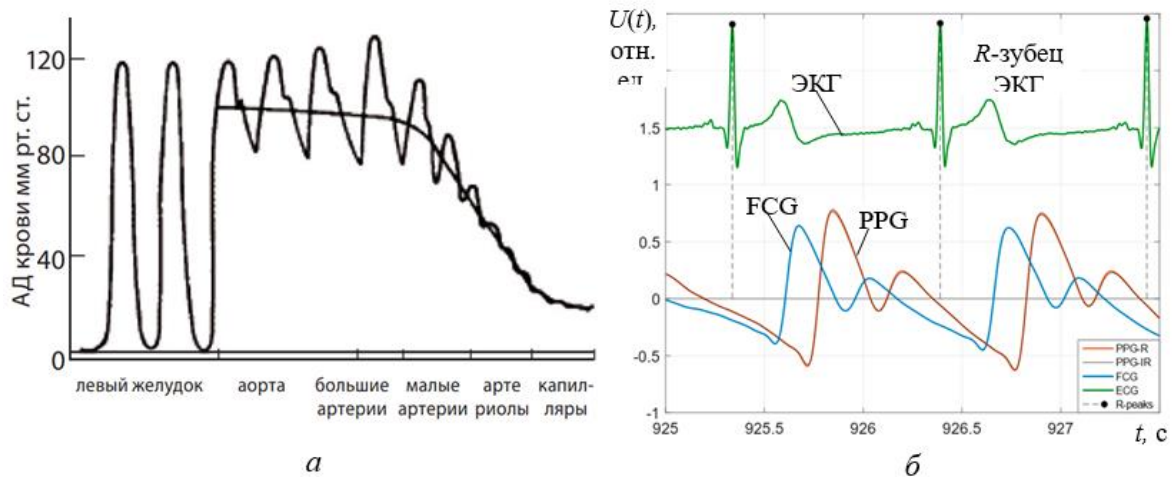
Несмотря на важность оценки состояния сосудов микроциркуляторного русла, они в настоящее время менее всего доступны для оперативного обследования из-за отсутствия соответствующих неинвазивных методик и аппаратуры. Ультразвуковые методы обследования для них неприменимы, поэтому перспективно применить оптические методы [1, 2], основанные на диффузном отражении светового излучения (видимого и ближнего инфракрасного диапазона спектра) живыми пульсирующими тканями конечностей.

КОМПЛЕКС ДЛЯ ОЦЕНКИ СОСТОЯНИЯ АРТЕРИАЛЬНЫХ СОСУДОВ

При диагностике артериальных сосудов необходимо оценивать несколько параметров, первым из которых является скорость распространения пульсовой волны. Для магистральных артериальных сосудов она определяется по задержкам между силовыми кардиоплетизмограммами (FCG) на сонной и бедренной артериях и R-зубцом электрокардиограммы

(ЭКГ). Это позволяет, зная расстояние между точками измерений и разность задержек FCG и ЭКГ, рассчитать СРПВ.

Для малых артериальных сосудов микроциркуляторного русла прямое получение FCG невозможно. Поэтому силовую кардиоплетизмограмму можно регистрировать, создавая стандартное натяжение в кольцевом направлении манжетки, надеваемой на палец конечности. Тогда, измеряя изменения силы давления, оказываемой тканью пальца на манжетку при сокращениях сердца, можно регистрировать максимум расширения ткани пальца при поступлении очередной порции крови за счет повышения ее давления в процессе распространения пульсовой волны. Действительно, как видно из временных зависимостей, позаимствованных из [3] и [4] и представленных на рис. 1, наблюдается задержка между R-зубцом ЭКГ и максимумом FCG(t), регистрируемой на дистальной фаланге пальца руки. Причем фотоплетизмограмма $PPG(t)$, которая одновременно регистрируется в этой же точке пальца с помощью модуля пульсоксиметра MAX30102, отстает от FCG(t).



Распределение артериального давления крови (АД) по сосудам [3] (а) и временные зависимости потенциала II-го отведения электрокардиограммы (ЭКГ), силовой кардиоплетизмограммы (FCG) и фотоплетизмограммы (PPG) дистальной фаланги пальца руки (б) [4]

Предположительно задержка между FCG(t) и PPG(t) отражает интервал времени, необходимый на заполнение артериального бассейна микроциркуляторного русла в исследуемой точке тела новой порцией крови за счет перепада периферийного давления крови, вызываемого дошедшей до этой точки пульсовой волной [2]. При периодическом заполнении артериол новой порцией крови возникает пульсирующее расширение диаметра внутреннего просвета артериол $D_{ao}(t)$, что в свою очередь увеличивает поглощения светового излучения кровью и формирует фотоплетизмограмму.

При ранней диагностике важна оценка резерва расширения площади внутреннего просвета артериол (который определяет диастолическое давление), поскольку при наличии этого резерва возможна медикаментозная коррекция тонуса артериол. Подход к решению этой проблемы на основе регистрации спектрально-временных профилей диффузного отражения ткани в исследуемой точке тела подробно описан в [4].

Таким образом для оценки состояния артериальных сосудов следует привлекать следующие диагностические признаки:

- скорость пульсовой волны в магистральных артериальных сосудах, что позволяет судить об их жесткости;
- временную задержку фотоплетизмограмм на пальцах руки и ноги относительно силовой плетизмограммы, что позволяет косвенно оценивать тонус артериол и скорость кровотока в микроциркуляторном русле.
- усредненный диаметр внутреннего просвета артериол пальцев рук и ног;
- коэффициент растяжения стенок артериол в кольцевом направлении.

Возможности современной спектральной, оптоэлектронной и компьютерной техники позволяют проводить такие неинвазивные измерения и оценивать имеющийся резерв расширения артериол.

В докладе приводится функциональная схема компьютеризированного комплекса, содержащего набор необходимых сенсоров (миниатюрный спектрофотометр C12880MA Hamamatsu, датчики FCG, PPG) и микроконтроллер для управления их работой, реализующего накопление регистрируемых данных, а также алгоритмы их обработки с выводом на экран получаемых параметров и временных зависимостей. Делается вывод о перспективности его применения при диспансеризации населения.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Possibilities of Diffuse Reflectance Spectroscopy in Determining and Operational Control of the Optical Properties of Finely Dispersed Scattering Media / Hotra O., Firago V., Shuliko K. [et al.] // Electronics. 2023. Vol. 12, № 13, 2893.
2. *Фираго В.А.* Оценка жесткости малых артериальных сосудов поверхностных биотканей по их спектрально-временным профилям диффузного отражения светового излучения // Журнал прикладной спектроскопии. 2024. Т. 91, № 1. С. 107–123.
3. Распространение пульсовой волны по малым сосудам: Результаты измерений и подходы к моделированию / Кубарко А.И., Мансуров В.А., Светличный А.Д. [и др.] // Неотложная кардиология и кардиоваскулярные риски. 2020. Т. 4, № 2. С. 1037–1044.
4. Multimodal Finger PulseWave Sensing: Comparison of Forcecardiography and Photoplethysmography Sensors / Andreozzi E., Sabbadini, R., Centracchio J. [et al.] // Sensors. 2022. Vol. 22, 7566.

ВЕБ-ПРИЛОЖЕНИЕ ДЛЯ АНАЛИЗА ДАННЫХ ФЛУОРЕСЦЕНТНОЙ СПЕКТРОСКОПИИ

С. Ю. Хоменко, А.В. Дигрис

*Белорусский государственный университет, Минск, Беларусь,
E-mail: rct.homenkoSY@bsu.by*

В статье представлено веб-приложение, предназначенное для удаленного анализа данных флуоресцентной спектроскопии. Приложение построено на основе архитектуры клиент-сервер с применением стека современных технологий: платформы Java и фреймворка ReactJS. Обмен данными между клиентской и серверной частями приложения выполняется с использованием разработанного REST API. Созданное веб-приложение позволяет широкому кругу пользователей выполнять анализ данных, измеренных методами однофотонного счета и флуоресцентной флукуационной спектроскопии, получать оценки параметров, характеризующих исследуемых образцов, а также контролировать качество полученных результатов.

Ключевые слова: флуоресцентная спектроскопия; анализ данных; метод максимального правдоподобия; веб-приложение.

Введение. Экспериментальные методы флуоресцентной спектроскопии неизменно играют весомую роль при изучении сложных молекулярных систем [1]. Для получения информации об исследуемой системе на основе данных, измеряемых этими методами, требуется разработка специального программного обеспечения (ПО). Помимо реализации набора математических алгоритмов, предназначенных для проведения анализа [2, 3], современные требования к разработке такого ПО включают необходимость обеспечить обработку данных, полученных различными экспериментальными методами, а также сделать разработанное ПО удобным для использования и доступным широкому кругу исследователей. Как следствие, клиент-серверное ПО для обработки данных флуоресцентной спектроскопии, реализованное в виде веб-приложения, становится все более актуальным, поскольку оно является максимально доступным за счет использования сети Интернет, позволяет использовать облачные ресурсы, распределять вычислительную нагрузку между сервером и клиентом, что улучшает производительность и эффективность работы.

Целью данной работы является разработка веб-приложения с пользовательским интерфейсом, предоставляющего возможность быстрого и удобного анализа данных флуоресцентных экспериментов. Наиболее подходящим вариантом такого интерфейса является веб-интерфейс, позволяющий пользователю выполнять анализ посредством веб-браузера с устройств, использующих различные аппаратно-программные платформы без установки дополнительного программного обеспечения.

Веб-приложение. Структурная схема созданного веб-приложения приведена на рис. 1.

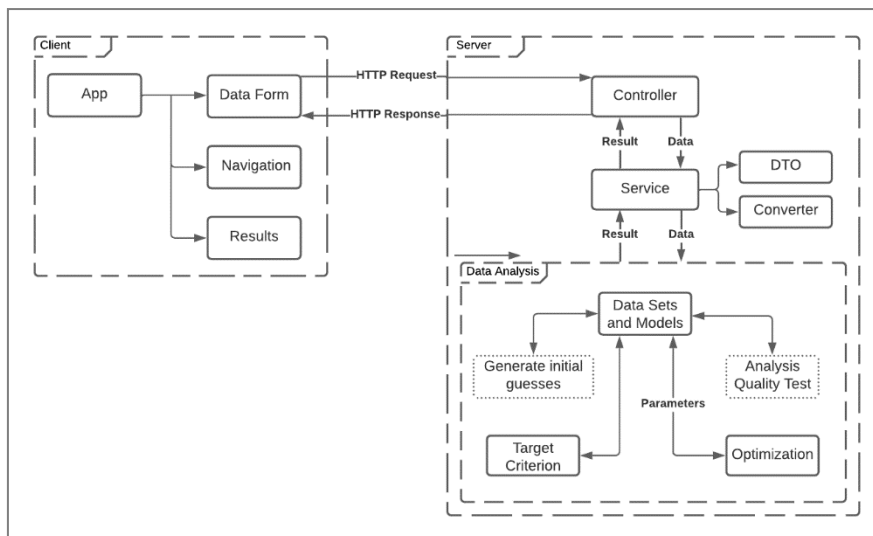


Рис. 1. Структурная схема веб-приложения для анализа данных флуоресцентной спектроскопии

Веб-приложение состоит из серверной и клиентской частей. Серверная часть реализована на базе платформы Java с использованием модулей фреймворка Spring [4] и представляет собой RESTful Spring Boot веб-службу. Клиентская часть реализована на языке JavaScript и состоит из компонент, написанных с использованием фреймворка ReactJS [5].

Для структурной организации приложения была выбрана трёхуровневая архитектура, в рамках которой приложение разделяется на слой представления, слой бизнес-логики и слой доступа к данным.

Слой представления отвечает за отображение данных пользователю, получение введенной им информации, и передачу ее в слой бизнес-логики для обработки. Слой представления построен с использованием архитектурного шаблона Model-View-Controller (MVC), который добавляет в структуру приложения три компонента, каждый из которых выполняет отдельную функцию. Model представлена в виде Data Transfer Object (DTO), которые определяют формат данных, передаваемых между компонентами веб-приложения. View является отдельным клиентским приложением, взаимодействующим с веб-службой через HTTP запросы в соответствии с разработанным REST API. Controller представлен классами, которые обрабатывают запросы и передают данные из них в бизнес-логику приложения.

Слой бизнес-логики отвечает за обработку запросов от представления, упаковку и распаковку DTO, валидацию и принятие решений касательно результатов работы с данными. Взаимодействие слоя представления с этим слоем происходит через контроллеры.

Слой доступа к данным представлен блоком их анализа, который отвечает за генерацию данных, которые впоследствии получит пользователь. Анализ данных представляет собой итерационную процедуру, основанную на методе максимального правдоподобия [2]. Текущая версия приложения позволяет обрабатывать кривые затухания флуоресценции, полученные методом однофотонного счета, с использованием много-экспоненциальной модели, либо автокорреляционную функцию потока фотонов, измеренного методом флуоресцентной флукуационной спектроскопии, с использованием модели свободной трехмерной диффузии [3].

Клиентская часть веб-приложения состоит из четырех основных компонент (смотри рис. 1), написанных с использованием фреймворка ReactJS. Компонент App является главным, определяет общую структуру веб-интерфейса и обеспечивает отрисовку компонентов Navigation, DataForm и Results, а также передачу данных между ними. Компонент Navigation содержит кнопки для выбора типа анализируемых данных в зависимости от метода измерений, которым они были получены. В соответствии с выбранным типом данных компонент DataForm предоставляет формы, используя которые пользователь может загрузить исходные данные для анализа в виде текстовых файлов, задать настройки модели, выбранной для анализа и, при необходимости, указать начальные приближения оцениваемых параметров. После нажатия на кнопку «Отправить», внесенная пользователем информация упаковывается в объект и с помощью http-клиента axios отправляется на серверную часть, где происходит ее обработка. Полученные в результате анализа результаты отправляются в компонент Results. Данный компонент выводит на экран графики экспериментальной и теоретической кривых, позволяющие оценить качество анализа визуально. Дополнительно отображаются полученные в результате анализа оценки параметров модели и значение критерия хи-квадрат.

Использование многослойной структуры с четко определенным функционалом отдельных слоев и набора компонент, решающих отдельные подзадачи в клиентской части, делает созданное веб-приложение гибким и доступным для дальнейшего расширения. Обмен данными между клиентской и серверной частями в стандартном формате JSON посредством http запросов обеспечивает возможность независимой разработки этих частей веб-приложения. Применение стека технологий Java и JavaScript позволяет развернуть веб-приложение на основных современных серверных платформах, в том числе под управлением ОС Linux.

Пример внешнего вида созданного веб-интерфейса, отображающего данные, полученные в результате анализа, приведен на рис. 2.

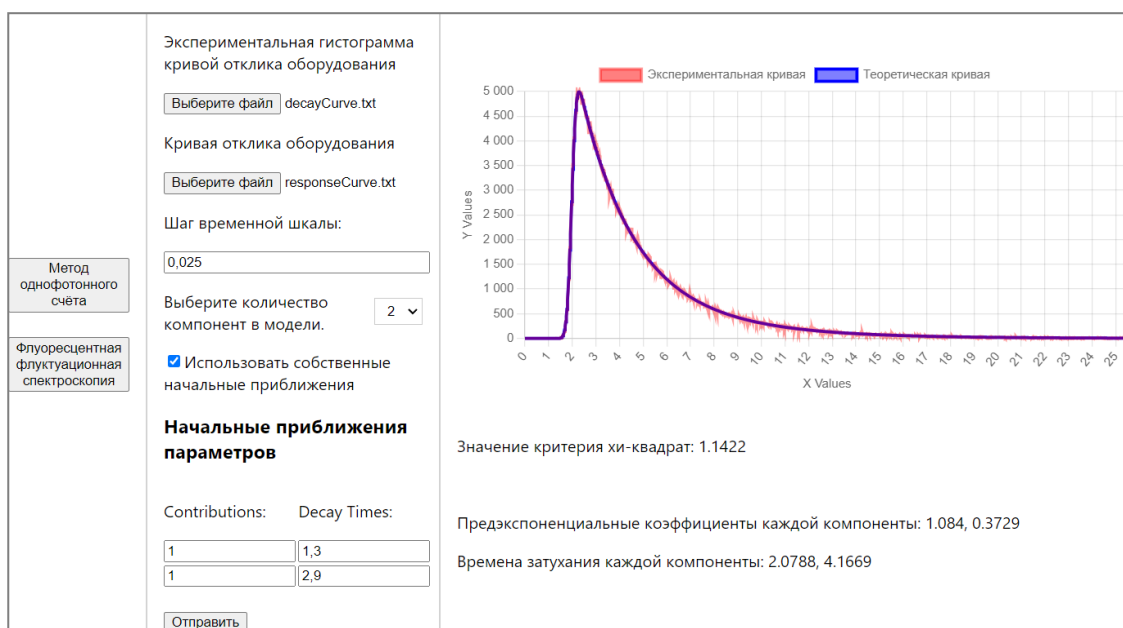


Рис. 2. Внешний вид интерфейса веб-приложения

Выводы. Созданное веб-приложение является широкодоступным инструментом для анализа данных, полученных различными экспериментальными методами флуоресцентной спектроскопии. Пользовательский интерфейс разработанного веб-приложения позволяет проводить анализ данных флуоресцентных экспериментов и получать в удобном виде информацию о качестве проведенного анализа. Разработанная структура веб-приложения позволяет наращивать его функционал добавлением новых моделей, средств оценки качества анализа и возможности анализировать данные, полученные другими экспериментальными методами флуоресцентной спектроскопии.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Lakowicz J. R. Principles of Fluorescence Spectroscopy // Springer; 3rd ed., 2006. P. 954.
2. Digris A. V., Novikov E. G., Skakun V. V., Apanasovich V. V. Global Analysis of Time-Resolved Fluorescence Data // In book Fluorescence Spectroscopy and Microscopy: Methods and Protocols: Methods in Molecular Biology, Springer Protocols, Yves Engelborghs and Antonie J.W.G. Visser (eds.). Springer Science+Business Media, LLC. 2014. V. 1076. P. 257-277.
3. Skakun V. V., Digris A. V., Apanasovich V. V. Global Analysis of Autocorrelation Functions and Photon Counting Distributions in Fluorescence Fluctuation Spectroscopy // In book Fluorescence Spectroscopy and Microscopy: Methods and Protocols: Methods in Molecular Biology, Springer Protocols, Yves Engelborghs and Antonie J.W.G. Visser (eds.). Springer Science+Business Media, LLC. 2014. V. 1076. P. 719-741.
4. Raman R. CSP, Dewayly L. Building RESTful Web Services with Spring // Packt Publishing; 2nd ed., 2018. P. 228.
5. Salcescu C. Functional React // Independently published; 2nd ed., 2020. P. 158.

Научное издание

**Компьютерные технологии
и анализ данных
(CTDA'2024)**

**Материалы
IV Международной научно-практической конференции**

Минск, 25–26 апреля 2024 г.

В авторской редакции

На русском и английском языках

Ответственный за выпуск *В. В. Скакун*