

ОГЛАВЛЕНИЕ

Часть 1. Тематический выпуск "НОМО LUDENS", II

Т. Р. Арсланов

ГЕНЕРАТИВНЫЕ МЕТОДЫ ДЛЯ СОЗДАНИЯ АДАПТИВНЫХ
ИГРАБЕЛЬНЫХ ПЕРСОНАЖЕЙ В ИГРАХ-СЕРВИСАХ 468–483

И. Е. Плотников, Д. И. Костюк

ПРЕДСТАВЛЕНИЕ ТРЁХМЕРНЫХ ОБЪЕКТОВ ДЛЯ ЛОГИЧЕСКИХ
ПРЕОБРАЗОВАНИЙ В РЕАЛЬНОМ ВРЕМЕНИ 484–505

А. М. Примаченко, М. Р. Хафизов

РАЗРАБОТКА СИСТЕМЫ ВИЗУАЛЬНОГО ВОСПРИЯТИЯ
ИГРОВЫХ АГЕНТОВ В ВИДЕОИГРАХ 506–531

A. de J. Parada Figueroa, O. Correa Madrigal

THE EDUCATIONAL VIDEO GAME FOR THE PROMOTION OF
TRADITIONAL CUBAN MUSIC AMONG CHILDREN AND TEENAGERS 531–543

М. В. Шматко, А. А. Кутузов, Л. Р. Пономарев

ИССЛЕДОВАНИЕ ИГРОВОГО ОПЫТА И ФОРМИРОВАНИЕ
КОГНИТИВНОЙ МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЕЙ
С НОВОЙ ВИДЕОИГРОЙ 544–572

Е. И. Ярko

СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК ДЛЯ ДЕТЕКТИРОВАНИЯ ПОЗЫ
ЧЕЛОВЕКА В УСЛОВИЯХ РАБОТЫ НА МОБИЛЬНЫХ УСТРОЙСТВАХ 573–600

Часть 2. Оригинальные статьи

М. В. Бобырь, Б. А. Бондаренко ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ НЕЙРО-НЕЧЕТКОГО ПОДХОДА	601–621
В. И. Гурьянов, А. М. Елизаров ПРОГРАММНЫЙ МОДУЛЬ ФОРМИРОВАНИЯ ЦИФРОВОГО МАТЕМАТИЧЕСКОГО ПРОСТРАНСТВА НА ОСНОВЕ ГРАФОВ ЗНАНИЙ	622–639
Ю. С. Ефимов ОПТИМИЗАЦИЯ АЛГОРИТМОВ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ C++ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МНОГОПОТОЧНОСТИ	640–653
И. И. Кузнецов, О. П. Новиков, Д. Ю. Ильин МЕТОДИКА СРАВНЕНИЯ ПРОГРАММНЫХ РЕШЕНИЙ РАСПОЗНАВАНИЯ ТЕКСТОВ НАУЧНЫХ ПУБЛИКАЦИЙ ПО КАЧЕСТВУ ИЗВЛЕЧЕНИЯ МЕТАДААННЫХ	654–680
К. А. Мащенко СИГНАТУРНЫЕ МЕТОДЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ	681–700
Р. Р. Минюков, М. М. Абрамский ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА ТЕСТОВ	701–717

ГЕНЕРАТИВНЫЕ МЕТОДЫ ДЛЯ СОЗДАНИЯ АДАПТИВНЫХ ИГРАБЕЛЬНЫХ ПЕРСОНАЖЕЙ В ИГРАХ-СЕРВИСАХ

Т. Р. Арсланов^[0009-0002-1110-8503]

Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

timars-mail@yandex.ru

Аннотация

В условиях роста популярности игр-сервисов, требующих постоянного обновления контента для удержания игроков, актуальной задачей становится автоматизация создания адаптивных играбельных персонажей. Нами рассмотрены существующие подходы к генерации персонажей.

Текущие решения не предусматривают долгосрочную адаптацию под стиль игрока и зависят от ручного проектирования. Для устранения этого недостатка предложена трёхкомпонентная система, сочетающая моделирование действий игрока на основе реплеев¹, генерацию персонажей через комбинирование механик и балансировку параметров, а также автоматическую валидацию через симуляции для оценки баланса и соответствия игровому стилю конкретного человека.

Работа обобщает современные исследования, демонстрируя потенциал генеративных методов для снижения ресурсозатрат при разработке игр-сервисов. Результаты могут быть использованы для ускорения прототипирования и поддержки долгосрочной жизнеспособности игровых проектов.

Ключевые слова: *игры-сервисы, гейм-дизайн, игровые персонажи, видео-игры, процедурная генерация.*

¹ Реплей – запись игрового процесса, позволяющая осуществлять его дальнейшее воспроизведение. Может быть реализована как последовательная запись состояний игровых объектов, запись последовательности входящих команд с сохранением сида (от англ. seed) – начального значения генератора псевдослучайных чисел – для воспроизведения «случайных» операций или как запись экрана.

ВВЕДЕНИЕ

В последнее время в игровой индустрии большую популярность получила модель монетизации игры как сервиса или игры-сервисы. К концу 2023 года 95% студий разработчиков видеоигр работали над игрой-сервисом либо планировали её выпуск [1, 2].

От обычной игры игра-сервис отличается длительным жизненным циклом. Для такой бизнес-модели критически важно обеспечить высокое удержание игроков, чтобы они продолжали постоянно заходить в игру.

Такие возврат и удержание достигаются регулярным выпуском нового контента, привлекательного для игроков. Многие игры-сервисы сфокусированы на добавлении новых персонажей, ставя перед разработчиками задачу создания играбельных персонажей (т. е. таких, которыми может управлять игрок), сбалансированных в рамках существующей игровой экосистемы и при этом интересных людям с различными стилями игры. Дополнительную сложность такая задача представляет для сервисов разных жанров, которые уже имеют большое количество персонажей, в том числе: MOBA² (как пример, игра *League of Legends* – 168 персонажей), tower defense³ (*Arknights* – 337 персонажей), hero-shooter⁴ (*Rainbow 6: Siege* – 74 персонажа).

Стоит отметить, что классические подходы к созданию персонажей в основном опираются на ручные проектирование, прототипирование и тестирование, требуя значительных временных и ресурсных затрат. Разработка персонажа может занимать порядка 10 недель (гейм-дизайнеры работают параллельно с другими специалистами) [3].

² MOBA – Multiplayer Online Battle Arena, многопользовательская онлайн-боевая арена. Жанр, в котором две команды игроков, в которых каждый игрок управляет своим персонажем, стремятся уничтожить главное здание противника.

³ Tower defense – башенная защита. Жанр, в котором игроки стремятся помешать компьютерному противнику дойти до цели путём размещения башен (персонажей) и использования способностей.

⁴ Hero-shooter – геройская «стрелялка». Жанр, в котором основной игровой процесс представлен стрельбой от лица различных персонажей, часто обладающих уникальными способностями.

СУЩЕСТВУЮЩИЕ ПОДХОДЫ

Существует множество исследований, посвящённых как генерации игр, так и отдельных их составляющих. Например, в [4] описаны существующие классификации игровых механик, а также схема работы системы генерации игровых механик на основе классификации MGMT. В [5] представлена система Ludi, которая служит для описания и генерации новых настольных игр, а также для оценки их качества.

В [6] предложено полностью избавиться от различных персонажей. В разработанном прототипе жанра MOBA игроки получают персонажей с одинаковыми характеристиками и самостоятельно выбирают в начале игровой партии нужные способности персонажей из общего перечня. Таким образом, проблема создания персонажей заменяется проблемой разработки способностей.

Помимо самостоятельного выбора способностей, игроки получают персонализированный опыт благодаря системе адаптации персонажа (CAS – Character Adaptation System). В ходе игровой сессии CAS непрерывно отслеживает использование способностей и автоматически повышает соответствующие характеристики персонажа (например, интеллект при попадании магической атакой).

Такой подход действительно обеспечивает адаптацию игровых персонажей, однако адаптивность обеспечивается только внутри одной игровой сессии, что, с одной стороны, повышает гибкость в конкретном матче, с другой – не позволяет адаптировать персонажа под общий стиль игры конкретного человека. Как отмечено авторами [6], малое количество игроков в прототипе не позволяет сделать вывод о том, насколько предложенная система облегчает или усложняет разработку способностей и правку баланса по сравнению с существующими практиками создания персонажей.

На способностях персонажей также фокусируется работа [7], где предложено использовать эволюционные алгоритмы для создания новых способностей. Такая система была реализована в игре с боевой системой, аналогичной JRPG, – две партии из 4 персонажей, по 4 способности у каждого, противостоят друг другу в пошаговом режиме. Способности имеют два свойства – «польза» и «цена», которые применяются к цели и увеличивают/уменьшают какую-либо характери-

стику соответственно. Эти значения для каждой характеристики составляют генотип. Значение функции приспособленности зависит от количества использований способности игроками. Перед каждым боем игроки могут заменить свои способности на сгенерированные, а невыбранные способности уходят в архив, который используется для следующих генераций.

По результатам тестирования этой системы было обнаружено появление более разнообразных стратегий и сборок персонажей, чем можно было ожидать от классических архетипов RPG. Однако система генерации способностей была ограничена рамками одной игровой сессии 16 боёв в среднем и не сохраняла результаты в новой игре. Кроме того, было проведено всего 6 игровых сессий. Как отмечает автор указанной работы, создание постоянного сервера и общего пула способностей для генетического алгоритма, а также добавление к способностям новых механик должны привести к генерации способностей, более подходящих различным стилям игроков.

В [8] для решения проблемы малого количества игровых сессий для оценки сгенерированного контента использованы компьютерные симуляции. В разработанном файтинге Sol также были использованы эволюционные алгоритмы для создания персонажей. Схема работы этой системы представлена на рис. 1.

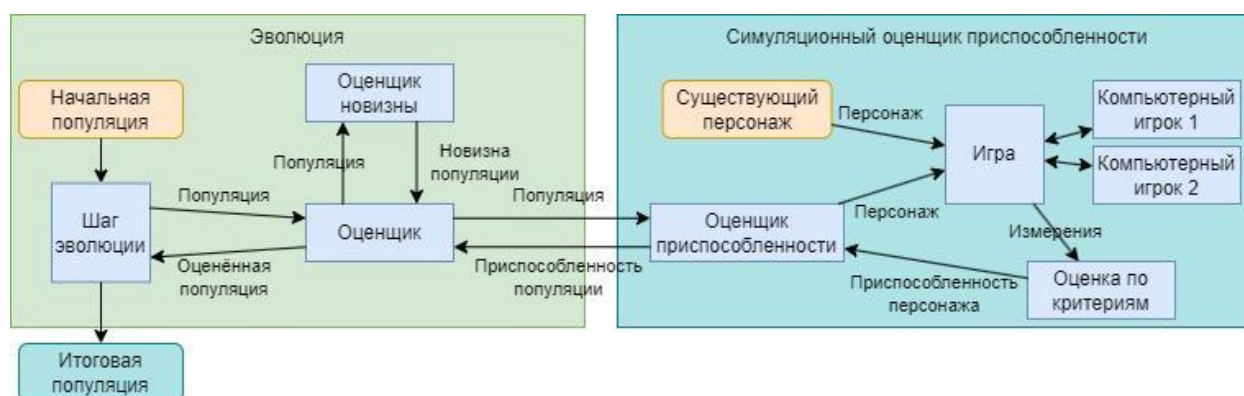


Рис. 1. Архитектура генератора [8]

Генотип составляют параметры персонажа и три способности. Создаваемые персонажи отправляются в симуляцию, где после серии игр оцениваются по 5 параметрам:

- сбалансированность матчей – процент побед персонажа;
- длительность матчей – средняя длительность матчей с персонажем;
- покрытие карты – среднее значение количества посещённых ячеек игровой карты (из 1400);
- сбалансированность персонажей – определяется средним соотношением количества использования способности персонажа к использованию всех трёх способностей того же персонажа за игру;
- смена лидерства – среднее число раз за игру, когда сменялся персонаж с лучшим состоянием.

Для реализации играющих агентов использовался подход, основанный на правилах. Для агентов было определено несколько правил: избегать пропасти, двигаться случайно, отступать от противника, подходить к противнику, избегать хитбоксов (от англ. hitbox – область поражения) – коллайдеров⁵ противника или снарядов, наносящих урон, или атаковать способностью (в т. ч. случайно).

Для исключения генерации похожих персонажей, помимо результатов симуляции в эволюции, также учитывалось евклидово расстояние до нескольких из ближайших соседей (обычно до 5) в пространстве генотипов персонажей.

В заключительной стадии исследования 6 сгенерированных и 6 разработанных человеком персонажей тестировались реальными игроками в игре против компьютера. После этого игрокам было предложено сравнить персонажей между собой по тому, насколько интересно им было играть за них. В результате сгенерированные персонажи получили несколько более высокие оценки, что подтверждает начальную гипотезу о возможности генерации «интересных» персонажей при помощи формальной оценки.

⁵ Коллайдер – область игрового пространства, пересечение объектов с которой фиксируется игрой как отдельное событие и вызывает соответствующий обработчик.

Генетические алгоритмы являются широко распространённым выбором для генерации игровых персонажей. В [9] показано их преимущество для генерации правдоподобных персонажей для футбольных менеджеров по сравнению со случайной генерацией и случайной генерацией с коррекциями, используемыми в рассматриваемом симуляторе.

Одной из существенных проблем персонализации чего-либо под конкретного игрока является отсутствие его индивидуальной модели. Ситуация усугубляется тем, что часто у разработчиков есть ограниченный набор реплеев игрока, недостаточный для обучения моделей. Решению этой проблемы на примере игры в реверси посвящена работа [10]. Алгоритм авторов проводит обучение модели в два этапа. С начала модель учится играть в игру в целом и не видит реплеев конкретного игрока. Затем подготовленная модель подстраивается на оставшихся записях (рис. 2).

Данная модель сравнивалась с нейронной сетью и случайным лесом. В ходе обучения модели схожесть с игроком не проверялась. Для проверки моделей использовались записи 6 ботов для игры реверси, обученных на разных алгоритмах.



Рис. 2. Схема имитационного моделирования отдельного игрока [10]

В качестве меры похожести бралась сумма совпадающих пар состояние – действие на тестовой выборке в 800 записей для каждого моделируемого бота. На обучающей выборке в 300 записей для каждого бота авторам удалось добиться от 33% до 59% похожести модели, что превышает результаты альтернативных моделей.

Несмотря на то что названный результат далёк от близкого копирования стратегии игрока, стоит отметить, что он достигнут на относительно малой обучающей выборке. Другим преимуществом этой модели является отсутствие эффекта

переобучения по сравнению с нейронными сетями при большом количестве итераций дообучения. В целом исследователи пришли к выводу, что универсальная обученная модель является хорошей точкой для создания модели игрока, т. к. стратегии разных игроков имеют общую составляющую.

Аналогично авторы [11] использовали предсказания следующих действий, но с учётом метрик Precision и Recall: сравнивали методы случайного леса и рекуррентных нейронных сетей с LSTM для имитации игроков в Starcraft. В качестве источника данных брались состояния игры каждые 3 секунды в реплеях 562 и 762 игр первого и второго игроков.

Результат исследования показал преимущество рекуррентных нейронных сетей для непустых действий. При этом показатели точности (Accuracy) у RNN находились в диапазоне 45–78%, что авторы называют недостаточным для удовлетворительной имитации действий игрока. С другой стороны, в 95% случаев действие игрока оказывалось в топ-3 предсказанных рекуррентной нейронной сетью, и, как предполагают авторы, при доработке выбора ответа модель сможет серьёзно улучшить свой результат.

АРХИТЕКТУРА ПРЕДЛАГАЕМОЙ СИСТЕМЫ

Несмотря на существующий прогресс в смежных вопросах, на данный момент отсутствует система, которая позволила бы генерировать интересные сбалансированные персонажи, подходящие для конкретного игрока в рамках не одной игровой сессии, а метаигры⁶.

В рамках разработки такой системы должны быть реализованы три подсистемы:

- подсистема, моделирующая действия игрока на основе записей его игрового процесса (реплеев);
- подсистема для автоматической генерации персонажей, подходящих игровому стилю игрока;

⁶ Метаигра – игровая и околоигровая деятельность игроков за пределами основного игрового цикла. Например, для игр, которые чётко разделены на партии, матчи и т. п., метаигра может включать анализ предыдущего опыта, прокачку персонажей, выполнение кумулятивных достижений, в которых считается прогресс за несколько партий, и т. д.

- подсистема автоматической валидации баланса и эффективности сгенерированного контента с помощью симуляций поведения игроков.

Для описания игрового стиля игрока предлагаются следующие элементы:

- время использования способностей;
- паттерны расположения персонажей на карте;
- стиль управления ресурсами;
- классовые предпочтения персонажей;
- агрессивный и оборонительный стиль игры;
- частота выбора персонажей;
- процент выигрышей при использовании персонажей различных классов.

Для анализа игрового стиля могут быть использованы реплеи, которые представляют собой временные ряды с действиями игрока.

Для имитации действий игрока в [12] предложено использовать параметрические деревья принятия решений. Такой подход представляется недостаточно универсальным. В рамках будущего исследования планируется использовать RL-агентов, при обучении которых с большим весом использовались бы партии игрока. Ранее в [13] RL-агенты применялись для тестирования уровней для игры жанра tower defense. Конкретная реализация будет выбираться в зависимости от результатов тестов, однако существуют исследования (например, [14]), которые показывают превосходство архитектуры Decision Transforming в задачах долгосрочного планирования перед другими подходами к обучению с подкреплением. С учётом того что в играх для достижения цели часто надо отклониться от локального оптимума, данный подход видится перспективным.

Состояния персонажей и других игровых сущностей при этом, так же как и используемые механики, могут быть описаны в рамках единого подхода, такого как FAST-GM [15]. В соответствии с ним представления сущностей содержат текущие значения атрибутов, ограничения на значения атрибутов, метку времени и вероятностные характеристики перехода в другие состояния. Помимо этого FAST-GM позволяет описывать также взаимодействия между различными сущностями, в том числе способности персонажей.

Для агента пространство состояний будет состоять из состояния персонажей, состояния способностей и сущностей на игровом поле, пространство действий, соответственно, – из выбора требуемого персонажа, указания его положения, использования способности и выбора цели персонажа.

Для генерации персонажей будет взят набор инструментов, его описывающих: комбинирование примитивных техник для создания способностей, подстройка числовых значений параметров персонажа и способностей, взаимное усиление способностей через дополнительные статусы. Кроме того, в виде ограничения в описании сущности по методологии FAST-GM будет использован набор балансовых ограничителей: стоимости, урона/исцеления в единицу времени, продолжительности эффектов контроля.

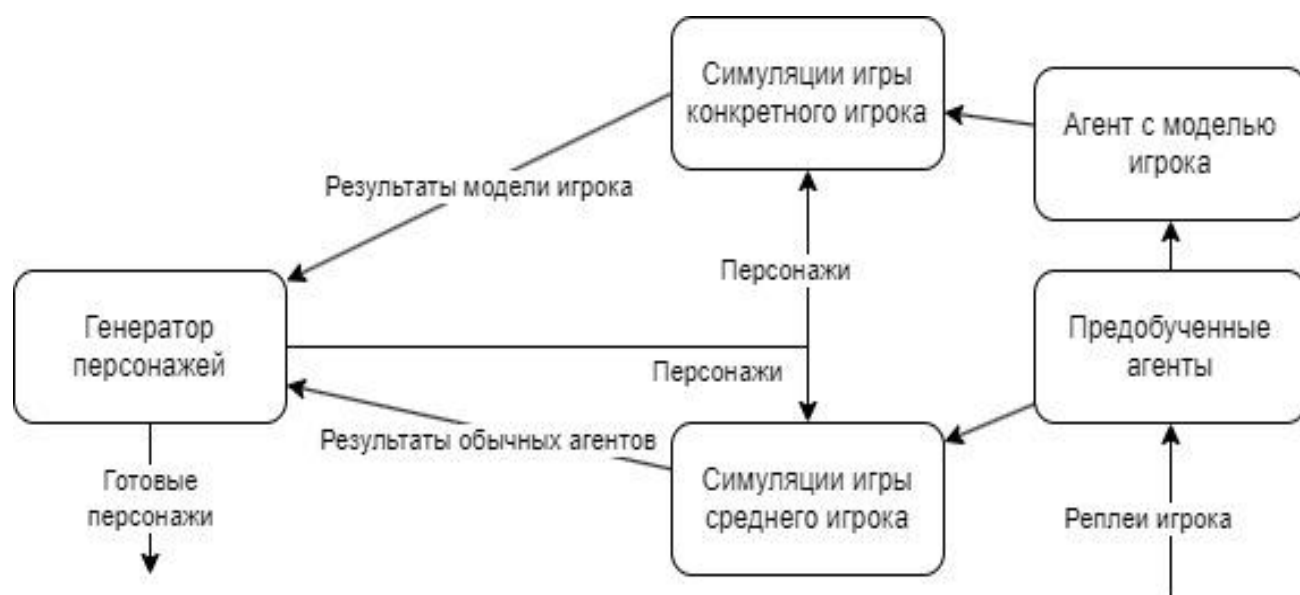


Рис. 3. Схема работы предлагаемой системы

После генерации следуют тестирование и оценка качества получившихся персонажей путём многократной игры обученных RL-агентов (рис. 3). Предполагается, что агенты, обученные на реплеях конкретного игрока, должны показывать результат не хуже, чем обычные агенты, чтобы игрок мог успешно использовать их в дальнейшем.

Помимо стандартных метрик, таких как процент побед с использованием персонажа, также будут реализованы метрики оценки схожести персонажей

со стилем игрока. В [12] для вычисления метрики схожести использованы матрица подобия для ресурсов игрока и её изменения с течением времени. В будущей работе предполагается учитывать соотношение урона, исцеления, наложения эффектов контроля и их изменение с течением времени.

ЗАКЛЮЧЕНИЕ

Рассмотрено текущее состояние исследований, связанных с процессами генерации персонажей и моделирования игроков для автоматического анализа контента.

Предложена система, учитывающая существующие результаты. Предполагено, что разрабатываемая система докажет возможность реализации автоматической генерации интересных адаптированных персонажей и будет обладать следующими параметрами:

- моделирование игроков с более чем 75% точностью для выбора персонажей, процента выигрыша персонажей, паттернов расположения;
- устойчивое определение игрового стиля игрока не более чем на 100 реплеях;
- устойчивая генерация персонажей, которые проходят оценку качества.

Успешная реализация предлагаемой системы откроет новые возможности для оптимизации процессов разработки и сопровождения игр-сервисов, значительно сократив время между концептуализацией и внедрением персонажей.

ЛУДОГРАФИЯ

Riot Games. (2009) *League of Legends*. [Видеоигра] [Загрузка] [Windows, MacOS]. Riot Games.

Hypergryph. (2020) [2019] *Arknights*. [Видеоигра] [Загрузка] [Android, iOS]. Yostar Games.

Ubisoft Montreal. (2015) *Tom Clancy's Rainbow Six Siege*. [Видеоигра] [Загрузка] [Windows, Xbox One, Xbox Series, Playstation 4, Playstation 5]. Ubisoft Entertainment.

Skjærseth E. и Vinje H. (2020) [2016] *Sol*. [Видеоигра] [Загрузка] [Windows]

Blizzard Entertainment. (1998) *Starcraft: Brood War*. [Видеоигра] [CD, Загрузка] [Windows, MacOS]. Blizzard Entertainment.

СПИСОК ЛИТЕРАТУРЫ

1. *Rousseau J.* Report: 95% of studios are working on or aim to release a live service game // GamesIndustry.biz. 2024. URL: <https://www.gamesindustry.biz/report-95-of-studios-are-working-on-or-aim-to-release-a-live-service-game>.
2. 2023 Game Development Report / Rendered Venture Capital, Griffin Gaming Partners. 2024. URL: <https://griffingp.com/2023-game-development-research/>.
3. *Canheti C, Andalo F., Vieira M.L.H.* Case Study: Game Character Creation Process // Advances in Human Factors in Wearable Technologies and Game Design: Advances in Intelligent Systems and Computing / T.Z. Ahram (Ed.). Cham: Springer International Publishing, 2019. Vol. 795. P. 343–354. https://doi.org/10.1007/978-3-319-94619-1_34.
4. *Шубин А.В., Кузурасова В.В.* Проектирование инструментария для создания игрового процесса через систематизацию игровых механик // Электронные библиотеки. 2024. Т. 27. № 5. С. 774–795. <https://doi.org/10.26907/1562-5419-2024-27-5-774-795>.
5. *Browne C, Maire F.* Evolutionary Game Design // IEEE Transactions on Computational Intelligence and AI in Games. 2010. Vol. 2, No. 1. P. 1–16. <https://doi.org/10.1109/TCIAIG.2010.2041928>.
6. *Cannizzo A., Ramírez E.* Towards Procedural Map and Character Generation for the MOBA Game Genre // Ingeniería y Ciencia. 2015. Vol. 11, No. 22. P. 95–119. <https://doi.org/10.17230/ingciencia.11.22.5>.
7. *Pantaleev A.* In Search of Patterns: Disrupting RPG Classes through Procedural Content Generation // Proceedings of the third workshop on Procedural Content Generation in Games FDG'12: International Conference on the Foundations of Digital Games. 2012. P. 1–5. <https://doi.org/10.1145/2538528.2538532>.
8. *Skjærseth E.H., Vinje H.* Evolutionary algorithms for generating interesting fighting game character mechanics: Master thesis. Trondheim: Norwegian University of Science and Technology, 2020. URL: <https://hdl.handle.net/11250/2777860>.
9. *No H.-S., Rhee D.-W.* A Study on The Game Character Creation Using Genetic Algorithm in Football Simulation Games // Journal of Korea Game Society. 2017. Vol. 17, No. 6. P. 129–137. <https://doi.org/10.7583/JKGS.2017.17.6.129>.

10. *Pan C.-F, Min X.-Y. et al.* Behavior imitation of individual board game players // *Applied Intelligence*. 2023. Vol. 53, No. 10. P. 11571–11585.
<https://doi.org/10.1007/s10489-022-04050-w>.
11. *Partla N. et al.* Player Imitation for Build Actions in a Real-Time Strategy Game // *AIIDE workshop on Artificial Intelligence for Strategy Games*. 2019.
12. *Kozik A., Machalewski T. et al.* Mimicking Playstyle by Adapting Parameterized Behavior Trees in RTS Games / *arXiv:2111.12144 [cs]*. 2021.
URL: <http://arxiv.org/abs/2111.12144>.
13. *Liu S., Chaoran L. et al.* Automatic generation of tower defense levels using PCG // *Proceedings of the 14th International Conference on the Foundations of Digital Games FDG'19: The Fourteenth International Conference on the Foundations of Digital Games*. San Luis Obispo California USA: ACM, 2019. P. 1–9.
<https://doi.org/10.1145/3337722.3337723>.
14. *Chen L., Lu K et al.* Decision Transformer: Reinforcement Learning via Sequence Modeling. *Decision Transformer* / *arXiv:2106.01345 [cs]*. 2021.
URL: <http://arxiv.org/abs/2106.01345>.
15. *Кузуркова В.В.* Формальный подход к пространственно-временному моделированию игровых систем // *Ученые записки Казанского университета. Серия Физико-математические науки*. 2024. Т. 166, №. 4. С. 532–554.
<https://doi.org/10.26907/2541-7746.2024.4.532-554>.

GENERATIVE METHODS FOR CREATING ADAPTIVE PLAYABLE CHARACTERS IN SERVICE GAMES

T. R. Arslanov^[0009-0002-1110-8503]

Kazan Federal University, Kazan, 420008, Russia

timars-mail@yandex.ru

Abstract

With the growing popularity of game services that require constant content updates to retain players, automating the generation of adaptive playable characters has become an urgent task. This article examines existing approaches to character generation, including evolutionary algorithms, and in-session adaptation systems. Current solutions are limited by their inability to provide sufficient long-term adaptation to individual player styles and their reliance on manual design.

To address these limitations, we propose a three-component system that integrates: player action modeling based on gameplay replays using reinforcement learning (RL) agents, character generation through combinatorial mechanics and parameter balancing, automatic validation via simulations to assess balance and alignment with a player's individual style.

This work synthesizes contemporary research, highlighting the potential of generative methods to reduce development costs for game services. The results could accelerate prototyping and enhance the long-term viability of live-service projects.

Keywords: *service games, game design, game characters, video game, procedural content generation.*

LUDOGRAPHY

Riot Games. (2009). *League of Legends* [Windows, MacOS]. Riot Games.
<https://www.leagueoflegends.com/>

Hypergryph. (2020). *Arknights* (Version Global) [Android, iOS]. Yostar.
<https://arknights.global>

Ubisoft Montreal. (2015). *Tom Clancy's Rainbow Six Siege* [Windows, Xbox One, Xbox Series, Playstation 4, Playstation 5]. Ubisoft Entertainment.
<https://www.ubisoft.com/en-us/game/rainbow-six/siege>

Skjærseth, E. H., & Vinje, H. (2020). *Sol* [Windows]. https://github.com/sol-ai/sol-lai_project

Blizzard Entertainment, & Saffire. (1998). *Starcraft: Brood War* [Windows, MacOS]. Blizzard Entertainment.

REFERENCES

1. Rousseau J. Report: 95% of studios are working on or aim to release a live service game // GamesIndustry.biz. 2024. URL: <https://www.gamesindustry.biz/report-95-of-studios-are-working-on-or-aim-to-release-a-live-service-game>.
2. 2023 Game Development Report / Rendered Venture Capital, Griffin Gaming Partners. 2024. URL: <https://griffingp.com/2023-game-development-research/>.
3. Canheti C, Andalo F., Vieira M.L.H. Case Study: Game Character Creation Process // Advances in Human Factors in Wearable Technologies and Game Design: Advances in Intelligent Systems and Computing / T.Z. Ahram (Ed.). Cham: Springer International Publishing, 2019. Vol. 795. P. 343–354. https://doi.org/10.1007/978-3-319-94619-1_34.
4. Shubin A.V., Kugurakova V.V. Designing a tool for creating gameplay through the systematization of game mechanics // Russian Digital Libraries Journal. Vol. 27, No. 5. P. 774–795. <https://doi.org/10.26907/1562-5419-2024-27-5-774-795>. <https://doi.org/10.26907/1562-5419-2024-27-5-774-795>.
5. Browne C, Maire F. Evolutionary Game Design // IEEE Transactions on Computational Intelligence and AI in Games. 2010. Vol. 2, No. 1. P. 1–16. <https://doi.org/10.1109/TCIAIG.2010.2041928>.
6. Cannizzo A., Ramírez E. Towards Procedural Map and Character Generation for the MOBA Game Genre // Ingeniería y Ciencia. 2015. Vol. 11, No. 22. P. 95–119. <https://doi.org/10.17230/ingciencia.11.22.5>.
7. Pantaleev A. In Search of Patterns: Disrupting RPG Classes through Procedural Content Generation // Proceedings of the third workshop on Procedural Content Generation in Games FDG'12: International Conference on the Foundations of Digital Games. 2012. P. 1–5. <https://doi.org/10.1145/2538528.2538532>.
8. Skjærseth E.H., Vinje H. Evolutionary algorithms for generating interesting fighting game character mechanics: Master thesis. Trondheim: Norwegian University of Science and Technology, 2020. URL: <https://hdl.handle.net/11250/2777860>.

9. *No H.-S., Rhee D.-W.* A Study on The Game Character Creation Using Genetic Algorithm in Football Simulation Games // Journal of Korea Game Society. 2017. Vol. 17, No. 6. P. 129–137. <https://doi.org/10.7583/JKGS.2017.17.6.129>.
 10. *Pan C.-F, Min X.-Y. et al.* Behavior imitation of individual board game players // Applied Intelligence. 2023. Vol. 53, No. 10. P. 11571–11585. <https://doi.org/10.1007/s10489-022-04050-w>.
 11. *Partla N. et al.* Player Imitation for Build Actions in a Real-Time Strategy Game // AIIDE workshop on Artificial Intelligence for Strategy Games. 2019.
 12. *Kozik A., Machalewski T. et al.* Mimicking Playstyle by Adapting Parameterized Behavior Trees in RTS Games / arXiv:2111.12144 [cs]. 2021. URL: <http://arxiv.org/abs/2111.12144>.
 13. *Liu S., Chaoran L. et al.* Automatic generation of tower defense levels using PCG // Proceedings of the 14th International Conference on the Foundations of Digital Games FDG'19: The Fourteenth International Conference on the Foundations of Digital Games. San Luis Obispo California USA: ACM, 2019. P. 1–9. <https://doi.org/10.1145/3337722.3337723>.
 14. *Chen L., Lu K et al.* Decision Transformer: Reinforcement Learning via Sequence Modeling. Decision Transformer / arXiv:2106.01345 [cs]. 2021. URL: <http://arxiv.org/abs/2106.01345>.
 15. *Kugurakova V.V.* A formal approach to spatio-temporal modeling of game systems // Uchenye Zapiski Kazanskogo Universiteta. Seriya Fiziko-Matematicheskie Nauki. 2024. Vol. 166, No. 4. P. 532–554. <https://doi.org/10.26907/2541-7746.2024.4.532-554>.
-

СВЕДЕНИЯ ОБ АВТОРЕ



АРСЛАНОВ Тимур Рузелевич – аспирант Института информационных технологий и интеллектуальных систем Казанского федерального университета. Направление исследований: фундаментальные аспекты создания видеоигр с высокой реиграбельностью.

Timur Ruzelevich ARSLANOV – Postgraduate student at the Institute of Information Technologies and Intelligent Systems of Kazan Federal University. Research area: fundamental aspects of creating video games with high replayability.

email: timars-mail@yandex.ru

ORCID: 0009-0002-1110-8503

Материал поступил в редакцию 25 января 2025 года

ПРЕДСТАВЛЕНИЕ ТРЁХМЕРНЫХ ОБЪЕКТОВ ДЛЯ ЛОГИЧЕСКИХ ПРЕОБРАЗОВАНИЙ В РЕАЛЬНОМ ВРЕМЕНИ

И. Е. Плотников¹ [0009-0004-2891-9913], Д. И. Костюк² [0000-0002-6542-6980]

^{1, 2}Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

¹ilaya.link@gmail.com, ²xdxnkx@gmail.com

Аннотация

Дан сравнительный анализ методов представления трёхмерных объектов для выполнения булевых операций в реальном времени в среде игрового движка Unity. Рассмотрены четыре основных подхода: полигональное представление на основе конструктивной твердотельной геометрии (CSG), функции знакового расстояния (SDF), воксельные методы и CAD-системы с представлением границ (B-Rep) и NURBS-поверхностями.

Проведено экспериментальное исследование производительности полигональных алгоритмов булевых операций и SDF-функций на основе реализации ray marching. Выявлено, что полигональные методы характеризуются высокими начальными затратами на построение системы, но обеспечивают стабильную производительность при длительных операциях и сохранение результатов преобразований. SDF-функции демонстрируют высокую скорость выполнения операций и гибкость в создании сглаженных переходов между объектами, однако ограничены в применении для долговременных задач из-за особенностей вычислительной модели.

Определены области эффективного применения каждого подхода: полигональные методы рекомендуются для задач, требующих точного геометрического контроля и интеграции с традиционными графическими конвейерами, в то время как SDF-функции оптимальны для процедурной генерации, многослойного рендеринга материалов и создания динамических визуальных эффектов. Результаты исследования могут быть использованы при разработке интерактивных симуляторов, игровых приложений и систем виртуальной реальности.

Ключевые слова: булевы операции, трёхмерное моделирование, конструктивная твердотельная геометрия, функции знакового расстояния, SDF-функции, Unity, реальное время, полигональные сетки, ray marching.

ВВЕДЕНИЕ

Булевы операции являются важным инструментом в трёхмерном моделировании, позволяя формировать новые геометрические объекты путём логического комбинирования базовых форм (см. рис. 1). Теоретические основы булевых операций над твердотельными объектами были заложены в [1], где впервые систематизированы принципы представления и обработки трёхмерной геометрии. Применение таких операций востребовано в различных сферах, включая проектирование изделий, архитектурное моделирование, создание интерактивного и игрового контента, а также в системах виртуальной и дополненной реальности.

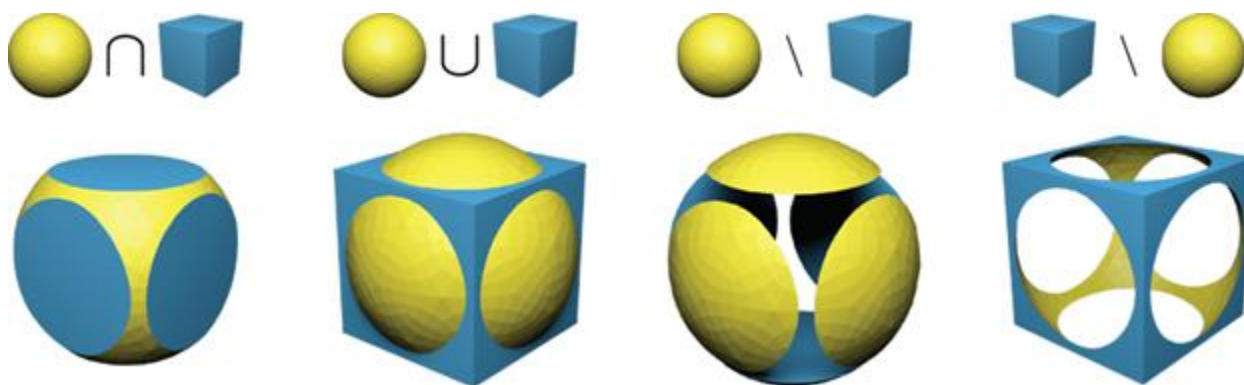


Рис. 1. Представление булевых операций для трёхмерных фигур

Особую актуальность приобретает задача выполнения логических операций в реальном времени, поскольку это открывает широкие возможности для динамической модификации объектов и интерактивного взаимодействия с виртуальной средой. Однако реализация таких операций предъявляет высокие требования к вычислительным ресурсам и производительности используемых алгоритмов. Эффективность выполнения логических операций напрямую зависит от выбранного способа представления трёхмерных объектов, алгоритмов обработки геометрии и особенностей программной платформы.

В настоящее время существуют различные подходы к описанию и представлению трёхмерных объектов, каждый из которых имеет свои преимущества

и ограничения при реализации булевых операций. Несмотря на наличие специализированных инструментов и библиотек, далеко не все существующие решения способны обеспечить требуемые производительность и качество выполнения операций в условиях реального времени.

Целью настоящего исследования были сравнительный анализ существующих методов и готовых реализаций для представления 3D-объектов и выполнения логических операций в реальном времени в среде игрового движка Unity и выработка рекомендаций по их эффективному использованию и оптимизации.

1. ОБЗОР СУЩЕСТВУЮЩИХ ПОДХОДОВ

1.1. Полигональное представление объектов

Полигональная геометрия остаётся наиболее распространённым способом представления трёхмерных объектов, особенно в игровых движках. Объекты описываются с помощью вершин, рёбер и полигонов, формирующих сетку поверхности. Булевы операции в таких моделях выполняются за счёт геометрического пересечения полигонов, классификации граней и перестройки меша (см. рис. 2) на основе логических условий (объединение, пересечение, разность).

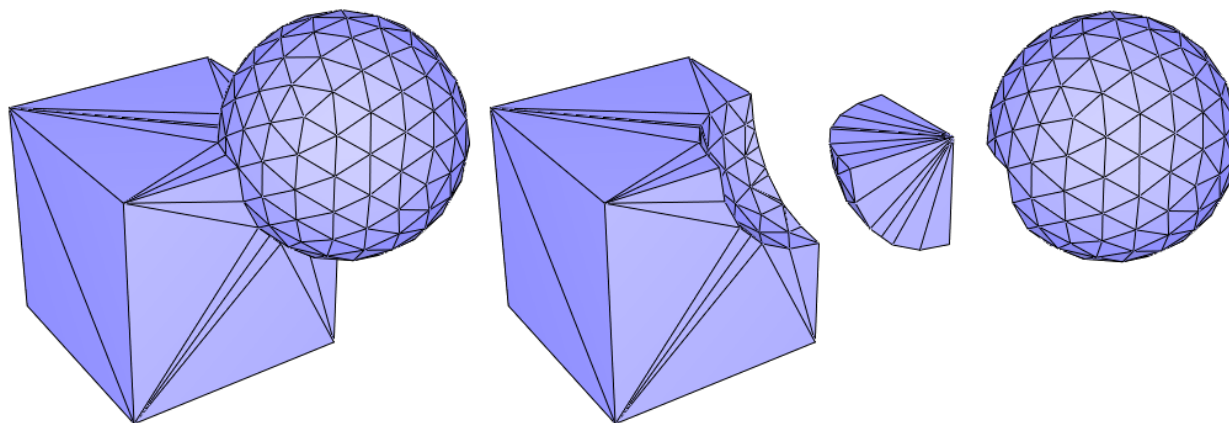


Рис. 2. Примеры операции вычитания над полигональными объектами

Преимущества полигонального представления объектов:

- возможность точной геометрии и контроля структуры модели;
- высокая совместимость с игровыми движками и стандартными средствами визуализации;
- поддержка анимации, освещения и текстурирования на уровне вершин и полигонов.

У полигонального представления имеются критичные недостатки:

- высокая чувствительность к качеству сетки – ошибки топологии, открытые грани;
- трудности с корректной обработкой самопересечений и пограничных случаев;
- высокая вычислительная сложность при перестроении полигональных сеток в реальном времени;
- необходимость в точном сопоставлении вершин, нормалей и UV-координат.

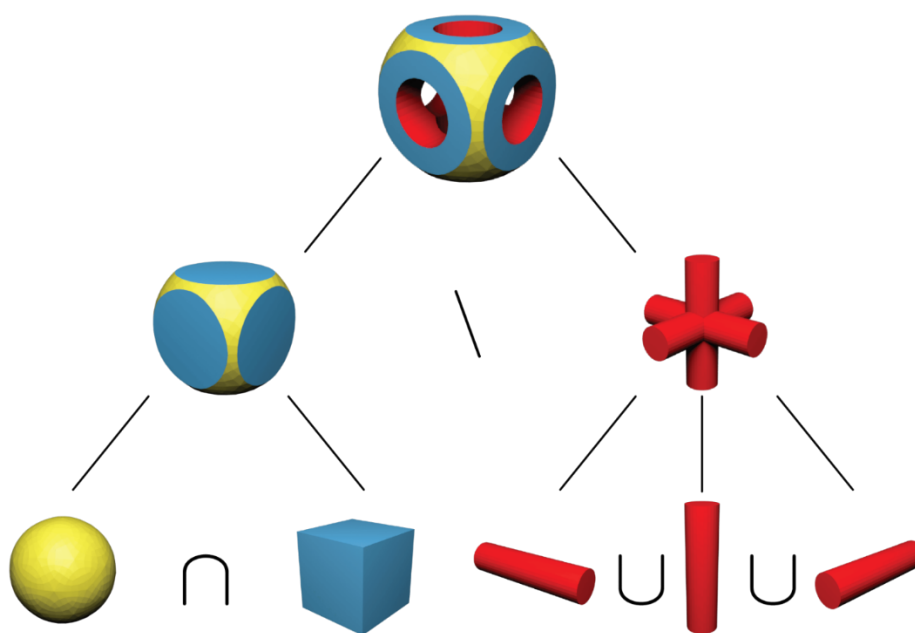


Рис. 3. Дерево булевых операций

Классическим примером реализации полигональной геометрии является конструктивная твердотельная геометрия (Constructive Solid Geometry, CSG), при которой булевы операции выражаются в виде дерева (см. рис. 3), где листьями выступают примитивы, а узлами – логические операции. Это метод Лейдлоу–Трампа–Хьюза, который был предложен в [2].

Итак, полигональные подходы, несмотря на широкую распространённость, предъявляют жёсткие требования к геометрии и требуют сложной обработки при выполнении булевых операций в реальном времени.

Математические основы конструктивной твердотельной геометрии и методы представления границ подробно рассмотрены в фундаментальной работе

[3]. Современные подходы к обеспечению робастности булевых операций [4] предлагают устойчивый алгоритм сегментации внутренних и внешних областей геометрических объектов.

1.2. Представление через поля расстояний

Функции знакового расстояния (Signed Distance Functions, SDF), другой часто используемый термин – «поля расстояний», представляют собой математический аппарат для описания трёхмерных объектов через скалярные поля, где каждая точка пространства характеризуется минимальным расстоянием до ближайшей поверхности объекта с учетом знака, указывающего положение точки относительно границы тела. Данный подход обеспечивает непрерывное и дифференцируемое представление геометрии, что критически важно для выполнения булевых операций в реальном времени без потери точности и с возможностью применения градиентных методов оптимизации. В отличие от дискретных представлений, таких как воксельные структуры или полигональные сетки, SDF сохраняют аналитическую точность на всех масштабах и позволяют эффективно реализовывать сложные топологические преобразования посредством алгебраических операций над скалярными функциями.

Концепция неявных поверхностей и методы их визуализации восходят к работе [5] по гипертекстурам. Метод трассировки сфер (Sphere tracing) как основной метод рендеринга SDF был формализован в [6]. Кроме того необходимо отметить всеобъемлющий обзор применений полей расстояний в компьютерной графике, который представлен в [7].

Этот подход предполагает описание формы объекта функцией расстояния до поверхности, что позволяет выполнять булевы операции в аналитическом виде: объединение реализуется как минимум двух значений, пересечение – максимум, а разность – как максимум первого и отрицания второго.

Отметим явные преимущества SDF-подхода:

- простота реализации булевых операций через математические выражения;
- хорошая масштабируемость при реализации на GPU;
- возможность использования сглаживающих функций для плавных переходов между объектами;

- поддержка процедурной геометрии и деформаций на лету.

Недостатки SDF-подхода:

- несовместимость со стандартными графическими конвейерами и необходимость применения объемных методов рендеринга (ray marching), исключающих использование аппаратно-ускоренной растеризации;
- ограниченная точность визуального отображения деталей при низком разрешении;
- сложность интеграции с физическим взаимодействием и коллизиями;
- требуется перепроектирование освещения и теней под объемное представление.

SDF-подход особенно привлекателен в задачах процедурной генерации и моделирования органических форм, но требует нетривиальной интеграции с классическим пайплайном визуализации. Дополнительным преимуществом является возможность применения сглаживающих функций (smooth-min) [8], создающих плавные переходы между формами (см. рис. 4). Это делает SDF-подход особенно привлекательным для визуально мягких и динамически изменяемых объектов [9].

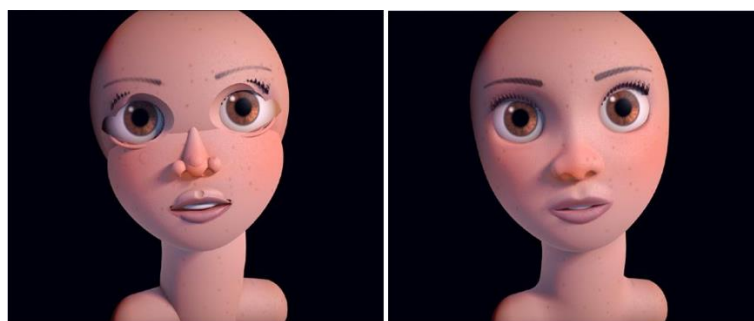


Рис. 4. Результаты проведения булевых функций [9]:

а) без сглаживающих функций; б) со сглаживающими функциями

Вместе с тем, как было отмечено ранее, визуализация таких моделей требует специальных алгоритмов, например ray marching [10], что ограничивает их применение в классическом рендеринге с полигональной сеткой.

1.3. Воксельные методы

Воксельные представления моделируют объекты как массив равномерных ячеек в трёхмерном пространстве. Булевы операции реализуются побитово, такой подход удобен для реализации разрушения, модификации или генерации сред в стиле игр Teardown или Minecraft.

Классическим алгоритмом визуализации воксельных данных является метод «марширующих кубов» (Marching Cubes) [11], обеспечивающий эффективное преобразование объемных данных в полигональные поверхности.

Преимущества воксельного представления:

- простота логических операций и устойчивость к топологическим ошибкам;
- естественная интеграция с физическим разрушением и симуляцией жидкостей;
- подходит для визуализации стилизованных миров (low-poly¹, cubic design²).

Недостатки этого подхода:

- высокая нагрузка на память при высокой детализации;
- необходимость последующей полигонализации для традиционного рендеринга;
- потери точности при преобразовании из других форматов;
- сложности с анимацией и точным управлением геометрией.

Воксельные подходы хорошо работают в сценариях, ограниченных по стилю или техническим условиям, но плохо масштабируются на реалистичную графику. Основными недостатками являются высокая нагрузка на память и ограничение по разрешению. Кроме того, детализация моделей ограничена размером вокселя, а реализация визуализации требует последующей полигонализации или специфического шейдерного рендеринга.

¹ Low-poly – стиль 3D-моделирования, характеризующийся использованием минимального количества полигонов для создания геометрических форм. Отличается упрощенными, угловатыми поверхностями и намеренно видимой полигональной структурой объектов.

² Cubic design — стилистический подход в 3D-графике, основанный на кубических и прямоугольных формах с четкими гранями и прямыми углами. Широко применяется в играх типа Minecraft и архитектурной визуализации для создания стилизованных блочных миров.

1.4. CAD-системы и инженерные решения

Булевы операции широко используются в инженерных CAD-системах (AutoCAD, SolidWorks, Fusion 360 и др.), где модели представлены в виде точных поверхностей и твёрдых тел, основанных на методе представления границ (boundary representation, B-Rep) и NURBS³. Эти системы обеспечивают высокую точность, контроль над параметризацией и удобство для последующего производства.

Однако CAD-подходы ориентированы преимущественно на офлайн-обработку и не адаптированы для выполнения логических операций в реальном времени. Кроме того, их форматы требуют конверсии при интеграции в игровые движки, что может приводить к потере точности или увеличению числа полигонов. Несмотря на это, принципы, используемые в CAD, легли в основу многих CSG-реализаций и алгоритмов булевых операций на уровне геометрических примитивов.

1.5. Сравнительный анализ

На основании проведенного анализа (табл. 1) можно заключить, что выбор подхода зависит от целей приложения.

³ NURBS (Non-Uniform Rational B-Splines) – математический метод представления кривых и поверхностей в компьютерной графике и CAD-системах, основанный на рациональных B-сплайнах с неравномерным распределением узловых точек. Обеспечивает точное описание как простых геометрических примитивов (окружности, эллипсы), так и сложных органических форм с высокой степенью гладкости и возможностью локального управления формой через контрольные точки.

Таблица 1. Сравнение рассмотренных подходов

Подход	Гибкость	Производительность	Визуальное качество	Сложность реализации	Работа в реальном времени
Полигональный	Средняя	Низкая	Высокое (при корректной сетке)	Высокая	Ограниченная
SDF	Высокая	Высокая (на GPU)	Высокое (smooth union)	Средняя	Хорошая
Воксели	Средняя	Средняя	Среднее	Средняя	Возможна
CAD-системы	Низкая	Низкая	Высокое (точные поверхности)	Высокая	Отсутствует

В условиях интерактивного применения и ограничений по производительности наибольший интерес представляют два направления: полигональные модели, как традиционный и точный способ описания геометрии, и функции знакового расстояния (SDF), как современный метод, обеспечивающий гибкость и высокую производительность при визуализации.

2. МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ

Для исследования был взят метод эксперимента: проследить за скоростью создания системы представления, реализованной самостоятельно, и сравнить результаты. В качестве кандидатов на тестирование были выбраны подходы полигональной сетки и SDF-функций.

Кроме того, помимо эмпирических исследований, был проведен анализ алгоритмов и систем для того, чтобы выявить нефункциональные свойства (не поддающиеся измерениям).

2.1. Используемые алгоритмы и теоретическое обоснование

Полигональный подход

Для реализации булевых операций над полигональными объектами была разработана модульная система-надстройка, обеспечивающая гибкое взаимодействие между компонентами геометрической модели и эффективную передачу данных между уровнями иерархии представления.

В качестве базового алгоритма выполнения булевых операций выбран метод Лейдлоу–Трампа–Хьюза [2], который обеспечивает точное пересечение полиэдральных объектов через классификацию результирующих граней на основе методов вычислительной геометрии. Выбор данного алгоритма обусловлен его открытой доступностью и возможностью последующей оптимизации для задач реального времени. Для повышения эффективности пространственных запросов были использованы принципы организации динамических иерархий ограничивающих объемов [12].

SDF-подход

Реализация функций знакового расстояния основана на библиотеке Unity Ray Marching [13], которая интегрирует современные достижения в области SDF-моделирования. Математические основы используемых функций расстояния базируются на работах Quilez [14] по трёхмерным SDF-примитивам и методам их композиции.

Ключевой особенностью реализации является поддержка сглаживающих операций, позволяющих создавать плавные переходы между геометрическими объектами при выполнении булевых операций. Для организации сложных композиций SDF-функций применяется древовидная структура на основе «ограничивающий параллелепипед, выровненный по осям» (axis-aligned bounding box, AABB), обеспечивающая эффективную пространственную локализацию вычислений [15].

Теоретическое обоснование SDF-подхода опирается на фундаментальные исследования [7] по применению полей расстояний в компьютерной графике, а также на современные методы процедурного моделирования с использованием неявных поверхностей.

2.2. Реализация подходов в среде Unity

Для реализации SDF-функций была использована открытая библиотека [13].

Для полигональных операций была разработана специализированная система представления геометрических данных, основанная на унифицированном интерфейсе доступа к топологическим связям.

Архитектура системы (см. рис. 5) построена на четырехуровневой иерархии компонентов:

Корневой узел (`ComplexMesh`) – содержит коллекцию полигонов и обеспечивает глобальный доступ к структуре модели;

- **Полигоны** – инкапсулируют три ребра и поддерживают обратную ссылку на корневой узел;
- **Рёбра** – связывают пары вершин и хранят ссылки на смежные полигоны;
- **Вершины** – представляют геометрические точки с информацией о принадлежности к рёбрам.

Центральным элементом архитектуры является универсальный метод `GetRelatedComponents<T>(bool isUpwards)`, обеспечивающий эффективную навигацию по иерархии представления. Параметр `isUpwards` определяет направление поиска: при значении `true` выполняется обращение к родительским компонентам (например, получение корневого узла от вершины), при `false` – к дочерним элементам (получение вершин полигона).

Построение топологической структуры осуществлялось методом `ConstructMeshFromBottom()` по принципу восходящей иерархии: последовательное создание вершин, формирование рёбер на основе треугольной информации `Unity Mesh`, объединение рёбер в полигоны и агрегация всех полигонов в единую коллекцию корневого узла.

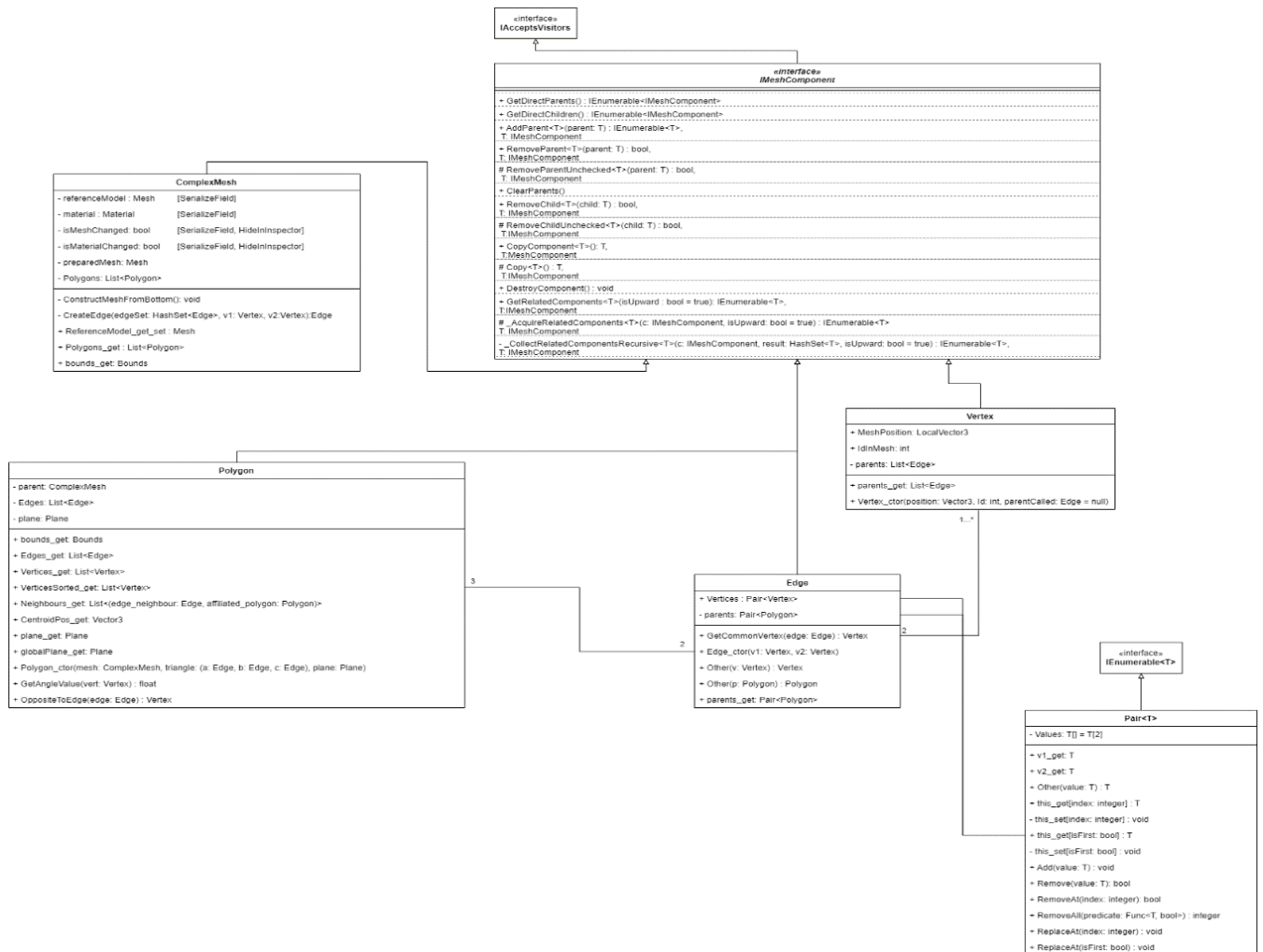


Рис. 5. Диаграмма классов разработанной системы

Каждый компонент содержит специализированную вычислительную информацию: полигоны хранят уравнения плоскостей, вершины – локальные координаты, что обеспечивает эффективность геометрических расчётов при выполнении булевых операций.

3. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

3.1. Описание тестовых моделей и среды

Для проведения сравнительного анализа был разработан единый программный модуль, включающий обе исследуемые библиотеки. В рамках эксперимента были созданы две идентичные по структуре сцены, в каждой из которых осуществлялось создание одного или нескольких объектов с использованием различных библиотек.

Измерение производительности выполнялось путём интеграции в исходный код функции таймера, осуществляющей точный замер временных интервалов, затрачиваемых на процесс построения трёхмерных моделей. Полученные временные показатели подвергались сравнительному анализу для определения эффективности каждой из исследуемых библиотек.

Выбор тестовых примитивов основан на стандартных практиках оценки производительности геометрических алгоритмов, где базовые формы (куб, сфера, цилиндр) обеспечивают воспроизводимость результатов и возможность сравнения с другими исследованиями [16].

3.2. Проведение экспериментов

В рамках исследования были выполнены тесты производительности для двух категорий геометрических объектов: полигональные сетки и SDF-функции.

Для анализа полигональных сеток использовались следующие тестовые модели:

1. Куб – базовый примитив, встроенный в среду разработки Unity;
2. Куб – модель, созданная в программе Autodesk Maya (для сравнения производительности при работе с встроенными примитивами и пользовательскими моделями, импортированными из внешних программных пакетов);
3. Цилиндр – стандартный геометрический примитив;
4. Сфера – сферический примитив;
5. Капсула – составной геометрический объект.

Поскольку для SDF-функций отсутствуют готовые примитивы, была проведена разработка соответствующих математических формул для создания аналогичных геометрических форм. Для каждого типа объекта были определены и реализованы математические выражения, позволяющие генерировать эквивалентные геометрические структуры посредством функций знакового расстояния [14].

Математические выражения для SDF-примитивов основаны на стандартных функциях расстояния, систематизированных в работах по неявному моделированию [6, 7].

3.3. Результаты экспериментов

Процедура измерения осуществлялась автоматически при назначении меша в соответствующее поле компонента. Встроенный в систему таймер фиксировал временные показатели выполнения операций. На основе полученных данных формировались графические представления результатов (рис. 5) для последующего анализа производительности обеих технологий.

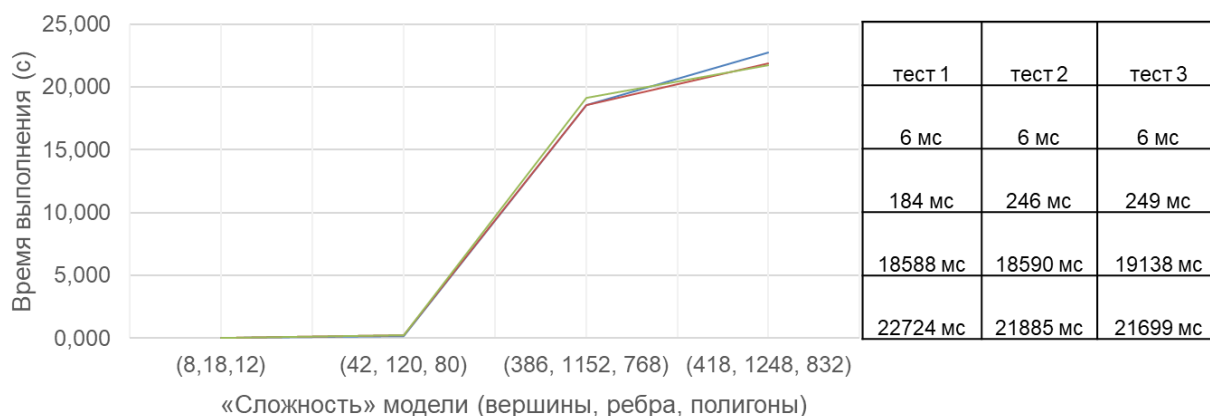


Рис. 5. Результаты построения системы полигональной сетки

Анализ производительности построения систем

Экспериментальные измерения показали существенные различия в характеристиках инициализации между исследуемыми подходами. Полигональные системы демонстрируют линейную зависимость времени построения от сложности геометрии, при этом время инициализации варьируется от 6 мс для базового куба до 21699 мс для сложных (см. рис. 5).

SDF-функции характеризуются константным временем инициализации, независимым от геометрической сложности описываемых объектов, поскольку представляют собой аналитические выражения без дискретных структур данных.

Анализ устойчивости результатов операций

Критическим различием между подходами является поведение системы при динамических изменениях геометрии. Полигональные представления обеспечивают сохранение результатов булевых операций: однажды выполненная операция сохраняет свой результат независимо от последующих трансформаций исходных объектов. Данное свойство обеспечивается фиксацией результирующей топологии в момент выполнения операции.

SDF-подход демонстрирует принципиально иное поведение: результат булевой операции вычисляется динамически для каждой пространственной точки на основе текущего положения операндов. Это приводит к тому, что перемещение одного из объектов после выполнения операции эффективно отменяет её результат, поскольку SDF-функция пересчитывает расстояния относительно новых позиций объектов. Выявленные ограничения SDF-подхода согласуются с теоретическими предсказаниями о природе функций знакового расстояния и их зависимости от пространственного контекста [7].

Влияние на долговременную производительность

Для решения проблемы непостоянства SDF-результатов требуется копирование состояния объектов на каждом кадре, что приводит к линейному росту вычислительных затрат со временем выполнения. Полигональные системы, напротив, демонстрируют стабильную производительность после завершения инициализации, поскольку результаты операций сохраняются в явном виде без необходимости пересчета.

Полигональные модели, несмотря на значительную вычислительную нагрузку при инициализации системы и выполнении алгоритма Лейдлоу–Трампа–Хьюза, обладают критически важным свойством устойчивого сохранения результатов операций. После завершения булевой операции система демонстрирует стабильную производительность без деградации со временем, поскольку результаты фиксируются в топологической структуре и не требуют пересчета при последующих трансформациях.

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Проведённые исследования выявили различные области применения рассматриваемых подходов в зависимости от временных требований задач. Полученные результаты согласуются с теоретическими предсказаниями о вычислительной сложности различных представлений трёхмерной геометрии [3, 6].

Традиционные CSG-алгоритмы с полигональными представлениями характеризуются высокими начальными затратами на построение и инициализацию системы. Однако данные методы поддаются существенной оптимизации и де-

монстрируют эффективность при выполнении длительных вычислительных операций, где время инициализации амортизируется за счёт продолжительности работы.

SDF-функции, напротив, обеспечивают высокую скорость выполнения операций изначально и не требуют сложной настройки системы. Тем не менее их применение ограничено краткосрочными задачами из-за накапливающихся вычислительных затрат при длительной работе.

Дополнительные возможности SDF-подхода проявляются в специализированных задачах многослойного рендеринга. В частности, при работе с объектами, состоящими из нескольких материальных слоев, SDF-функции позволяют эффективно определять принадлежность точек пересечения лучей к конкретным слоям геометрии. Это достигается путём вычисления знакового расстояния для каждой функции в точке пересечения и сопоставления результата с соответствующими текстурными данными, что особенно востребовано в задачах физического моделирования и визуализации сложных материалов.

5. ПЕРСПЕКТИВНЫЕ НАПРАВЛЕНИЯ ДАЛЬНЕЙШИХ ИССЛЕДОВАНИЙ

Результаты проведенного исследования указывают на несколько перспективных направлений для повышения эффективности систем представления трёхмерных объектов при выполнении булевых операций в реальном времени.

5.1. Оптимизация структур данных для полигональных представлений

Текущая реализация полигональной системы демонстрирует значительные возможности для оптимизации в области управления топологическими связями. В настоящее время система использует динамические запросы для построения иерархических отношений между геометрическими примитивами (полигон → рёбра → вершины), что приводит к избыточным вычислениям при каждом обращении к структуре данных.

Классические подходы к оптимизации топологических структур в полигональном моделировании [3] могут быть адаптированы для современных требований интерактивных приложений. Основным направлением оптимизации является внедрение системы кэширования топологических связей с предварительным

вычислением отношений родительских и дочерних элементов. Такой подход позволит исключить повторные обходы коллекций и устранить дублирование общих вершин и рёбер при обработке смежных полигонов. Однако реализация данного механизма в условиях покадрового обновления требует разработки эффективных алгоритмов инвалидации кэша и управления жизненным циклом кэшированных данных.

5.2. Гибридные подходы к представлению геометрии

Перспективным направлением является исследование гибридных методов, где полигональные представления используются для основной геометрии, а SDF-функции применяются для определения многослойных материальных свойств объектов. Такая архитектура позволит сочетать точность полигональных моделей с гибкостью SDF-подхода при реализации сложных материальных структур и процедурных текстур.

5.3. Адаптивные алгоритмы уровня детализации

Дальнейшие исследования должны быть направлены на разработку адаптивных систем управления уровнем детализации (Level of Detail), способных автоматически регулировать сложность геометрических представлений в зависимости от вычислительных ресурсов и требований к качеству визуализации.

ЗАКЛЮЧЕНИЕ

Проанализированные системы представления трёхмерных объектов демонстрируют значительный потенциал для создания реалистичных симуляций динамического изменения геометрии в виртуальном пространстве.

Применение операций над полигональными мешами обеспечивает получение физически корректных и устойчивых геометрических преобразований с возможностью создания новых топологических структур. Данный подход эффективен для выполнения конструктивных операций, таких как булево вычитание, экструзия и деформация сложных объектов.

Использование SDF-функций открывает дополнительные возможности для реализации многослойного представления материалов в интерактивных системах. Например, при моделировании процесса среза древесного объекта система способна корректно отображать внутреннюю структуру материала, демонстрируя

переход от коры к древесине с соответствующими текстурными характеристиками.

Практическое применение разработанных методов охватывает широкий спектр специализированных симуляторов: от медицинских тренажеров хирургических операций (см., например, [17, 18]) до интерактивных «песочниц». Кроме того, предложенные решения могут быть интегрированы в современные игровые движки в качестве библиотек для динамического редактирования импортированной геометрии, что особенно востребовано в процессе разработки уровней, где требуется адаптация базовых моделей под специфические дизайнерские требования.

СПИСОК ЛИТЕРАТУРЫ

1. *Requicha A.G.* Representations for rigid solids: Theory, methods, and systems // ACM Computing Surveys (CSUR). 1980. Vol. 12, No. 4. P. 437–464.
2. *Laidlaw D.H., Trumbore W.B., Hughes J.F.* Constructive solid geometry for polyhedral objects // Proceedings of the 13th annual conference on Computer graphics and interactive techniques. 1986. P. 161–170.
3. *Mäntylä M.* An introduction to solid modeling. Computer Science Press, Inc., 1987.
4. *Jacobson A., Kavan L., Sorkine-Hornung O.* Robust inside-outside segmentation using generalized winding numbers // ACM Transactions on Graphics (TOG). 2013. Vol. 32, No. 4. P. 1–12.
5. *Perlin K., Hoffert E.M.* Hypertexture // Proceedings of the 16th annual conference on Computer graphics and interactive techniques. 1989. P. 253–262.
6. *Hart J. C.* Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces // The Visual Computer. 1996. Vol. 12, No. 10. P. 527–545.
7. *Jones M.W., Baerentzen J.A., Sramek M.* 3D distance fields: A survey of techniques and applications // IEEE Transactions on visualization and Computer Graphics. 2006. Vol. 12, No. 4. P. 581–599.
8. *Quilez I.* Smooth Minimum. 2013. URL: <https://iquilezles.org/articles/smin>
9. *Quilez I.* Smooth Minimum on ShaderToy // ShaderToy. URL: <https://www.shadertoy.com/view/DIVcW1>
10. *Walczyk M.* Ray Marching. Personal Blog.

URL: <https://michaelwalczyk.com/blog-ray-marching.html>

11. Lorensen W.E., Cline H.E. Marching cubes: A high resolution 3D surface construction algorithm // *Seminal graphics: pioneering efforts that shaped the field*. 1998. P. 347–353.

12. Zhou K. et al. Data-parallel octrees for surface reconstruction // *IEEE Transactions on visualization and computer graphics*. 2010. Vol. 17, No. 5. P. 669–681.

13. Chou M.-L. Unity Ray Marching // GitHub.

URL: <https://github.com/TheAltenChou/unity-ray-marching>

14. Quilez I. 3D SDFs. 2013. URL: <https://iquilezles.org/articles/distfunctions/>

15. Catto E. Dynamic Bounding Volume Hierarchies // *Game Developers Conference*. 2019. URL: https://box2d.org/files/ErinCatto_DynamicBVH_GDC2019.pdf

16. Bernstein G., Fussell D. Fast, exact, linear booleans // *Computer Graphics Forum*. Oxford, UK: Blackwell Publishing Ltd, 2009. Vol. 28, No. 5. P. 1269–1278.

17. Шареева Р.А. и др. Подходы к проектированию виртуальных тренажёров хирургических операций // *Электронные библиотеки*. 2022. Т. 25, №. 5. С. 489–532.

18. Кузупракова В.В. и др. Virtual surgery system with realistic visual effects and haptic interaction // *Proc. of The International Conference on Artificial Life And Robotics*. 2017. С. P86–P89.

3D OBJECTS REPRESENTATION FOR REAL-TIME BOOLEAN OPERATIONS

I. E. Plotnikov¹ [0009-0004-2891-9913], D. I. Kostyuk² [0000-0002-6542-6980]

^{1, 2}Kazan Federal University, Kazan, 420008, Russia

¹ilaya.link@gmail.com, ²xdxnkx@gmail.com

Abstract

The paper presents a comparative analysis of methods for representing three-dimensional objects to perform real-time Boolean operations in the Unity game engine environment. Four main approaches are considered: polygonal representation based on constructive solid geometry (CSG), sign distance functions (SDF), voxel methods and CAD-systems with boundary representation (B-Rep) and NURBS-surfaces.

An experimental study of the performance of polygonal algorithms of Boolean operations and SDF functions based on ray marching implementation is carried out. It is revealed that polygonal methods are characterized by high initial system construction costs, but provide stable performance during long operations and preservation of transformation results. SDF functions demonstrate high speed of operations and flexibility in creating smooth transitions between objects, but are limited in application for long-term tasks due to the peculiarities of the computational model.

The areas of effective application of each approach are identified: polygonal methods are recommended for tasks requiring precise geometric control and integration with traditional graphics pipelines, while SDF functions are optimal for procedural generation, multilayer material rendering and creation of dynamic visual effects. The results of the study can be used in the development of interactive simulators, game applications and virtual reality systems.

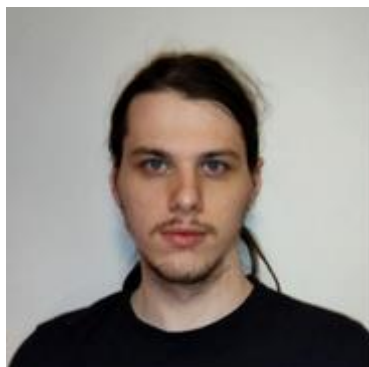
Keywords: *boolean operations, 3D modeling, constructive solid geometry, sign distance functions, SDF, Unity, real-time, polygonal meshes, ray marching.*

REFERENCES

1. Requicha A.G. Representations for rigid solids: Theory, methods, and systems // ACM Computing Surveys (CSUR). 1980. Vol. 12, No. 4. P. 437–464.
2. Laidlaw D.H., Trumbore W.B., Hughes J.F. Constructive solid geometry for polyhedral objects // Proceedings of the 13th annual conference on Computer graphics and interactive techniques. 1986. P. 161–170.
3. Mäntylä M. An introduction to solid modeling. Computer Science Press, Inc., 1987.
4. Jacobson A., Kavan L., Sorkine-Hornung O. Robust inside-outside segmentation using generalized winding numbers // ACM Transactions on Graphics (TOG). 2013. Vol. 32, No. 4. P. 1–12.
5. Perlin K., Hoffert E.M. Hypertexture // Proceedings of the 16th annual conference on Computer graphics and interactive techniques. 1989. P. 253–262.
6. Hart J. C. Sphere tracing: A geometric method for the antialiased ray tracing of implicit surfaces // The Visual Computer. 1996. Vol. 12, No. 10. P. 527–545.

7. Jones M.W., Baerentzen J.A., Sramek M. 3D distance fields: A survey of techniques and applications // IEEE Transactions on visualization and Computer Graphics. 2006. Vol. 12, No. 4. P. 581–599.
8. Quilez I. Smooth Minimum. 2013. URL: <https://iquilezles.org/articles/smin>
9. Quilez I. Smooth Minimum on ShaderToy // ShaderToy.
URL: <https://www.shadertoy.com/view/DIVcW1>
10. Walczyk M. Ray Marching. Personal Blog.
URL: <https://michaelwalczyk.com/blog-ray-marching.html>
11. Lorensen W.E., Cline H.E. Marching cubes: A high resolution 3D surface construction algorithm // Seminal graphics: pioneering efforts that shaped the field. 1998. P. 347–353.
12. Zhou K. et al. Data-parallel octrees for surface reconstruction // IEEE Transactions on visualization and computer graphics. 2010. Vol. 17, No. 5. P. 669–681.
13. Chou M.-L. Unity Ray Marching // GitHub.
URL: <https://github.com/TheAllenChou/unity-ray-marching>
14. Quilez I. 3D SDFs. 2013. URL: <https://iquilezles.org/articles/distfunctions/>
15. Catto E. Dynamic Bounding Volume Hierarchies // Game Developers Conference. 2019. URL: https://box2d.org/files/ErinCatto_DynamicBVH_GDC2019.pdf
16. Bernstein G., Fussell D. Fast, exact, linear booleans // Computer Graphics Forum. Oxford, UK: Blackwell Publishing Ltd, 2009. Vol. 28, No. 5. P. 1269–1278.
17. Sharaeva R. A. et al. Approaches to the design of virtual simulators of surgical operations // Russian Digital Libraries. 2022. Vol. 25, No. 5. P. 489–532.
18. Kugurakova V. et al. Virtual surgery system with realistic visual effects and haptic interaction // Proc. of The International Conference On Artificial Life And Robotics. 2017. P. P86–P89.

СВЕДЕНИЯ ОБ АВТОРАХ



ПЛОТНИКОВ Илья Евгеньевич – студент 2 курса магистратуры Института информационных технологий и интеллектуальных систем Казанского федерального университета. Сфера научных интересов: логические операции над твердотельными трехмерными объектами в реальном времени.

Ilya Evgenievich PLOTNIKOV – 2nd year master's student at the Institute of Information Technologies and Intelligent Systems of Kazan Federal University. Research interests: logical operations on solid three-dimensional objects in real time.

email: ilaya.link@gmail.com

ORCID: 0009-0004-2891-9913



КОСТЮК Даниил Иванович – старший преподаватель кафедры индустрии разработки видеоигр. Сфера научных интересов: процедурная генерация сцен, моделирование, рендеринг.

Daniil Ivanovich KOSTYUK – senior lecturer of the Department of Video Game Development Industry. Research interests include procedural scenes generation, modeling, rendering.

email: xdxnxkx@gmail.com

ORCID: 0000-0002-6542-6980

Материал поступил в редакцию 2 апреля 2025 года

РАЗРАБОТКА СИСТЕМЫ ВИЗУАЛЬНОГО ВОСПРИЯТИЯ ИГРОВЫХ АГЕНТОВ В ВИДЕОИГРАХ

А. М. Примаченко¹ [0009-0008-6396-2035], М. Р. Хафизов² [0000-0001-7275-9102]

^{1, 2}Казанский (Приволжский) федеральный университет, Казань, 420008, Россия

¹primachenko.artem@mail.ru, ²murkorp@gmail.com

Аннотация

Представлен алгоритм функционирования системы визуального восприятия для игровых агентов, реализованный в игровом движке Unity. Предложенный метод основан на сравнении изображений с двух камер, учитывающих сложные визуальные эффекты (освещение, тени, маскировку), и дополнен проверкой прямой видимости, учетом скорости движения объекта, и механикой постепенного обнаружения. Тестирование системы показало значительное повышение реалистичности обнаружения по сравнению с традиционными методами при сохранении производительности в пределах небольшой дополнительной нагрузки на процессор. Проведена оптимизация алгоритма с использованием Unity Job System и динамической активации камер. Проведен также анализ научной литературы по схожим решениям, выявлены их сильные и слабые стороны. Результаты могут быть применены в разработке видеоигр для создания реалистичного поведения неигровых персонажей, особенно в играх с элементами скрытности.

Ключевые слова: видеоигры, искусственный интеллект, система восприятия, NPC, неигровые персонажи, игровые агенты, стелс-механики, Unity, рендеринг, компьютерное зрение, оптимизация, гейм-дизайн.

ВВЕДЕНИЕ

Системы визуального восприятия неигровых персонажей (non-player character, NPC) в видеоиграх являются важным элементом, обеспечивающим реалистичность и интерактивность игрового процесса. Как отмечено в ряде работ (см., например, [1–7]), информация о видимости (или линии обзора) является ключевым элементом, позволяющим NPC строить пространственное представление о мире игры и принимать решения в реальном времени. Традиционные подходы, такие как проверка углов обзора и использование отражения лучей (или рейкастов, от англ. raycast) [8], обладают ограничениями, связанными с недостаточным учетом визуальных факторов, таких как освещение, тени и маскировка. Это приводит к неестественному поведению NPC, особенно в играх, использующих игровые механики невидимости («стелс-играх», от англ. stealth), где требуется высокая степень реализма [9].

Наиболее распространённым способом реализации системы зрения для неигровых персонажей является алгоритм, состоящий из двух шагов: проверка нахождения целевого объекта в определённой зоне в игровом мире, часто динамически изменяющейся в зависимости от положения персонажа и называемой полем зрения, и проверка на наличие препятствий между персонажем и целевым объектом. Хотя такой подход прост в реализации, он не учитывает множество факторов реального визуального восприятия, что снижает убедительность игрового опыта.

Целью проведенного исследования была разработка системы визуального восприятия, которая учитывает сложные визуальные эффекты, обеспечивает гибкость настройки и оптимизирована для реального времени. Объектом исследования выступают методы реализации визуального восприятия NPC, а предметом — алгоритм, основанный на сравнении изображений и дополнительных механиках.

1. СВЯЗАННЫЕ РАБОТЫ

Современные исследования в области систем визуального восприятия NPC предлагают разнообразные подходы.

Понятие базового рейкастинга (raycasting) для определения первого столкновения луча с объектом сцены было введено ещё в 1989 году [8].

В [10] представлен разработанный алгоритм, основанный на проверке нахождения объекта в поле зрения, представленного усеченной пирамидой, и использовании рейкаста для исключения препятствий. Этот метод включает проверку трех основных условий.

1. Ограничительная рамка (Bounding box) целевого объекта должна находиться внутри усечённой пирамиды поля зрения персонажа либо пересекать хотя бы одну из шести её граней (рис. 1).

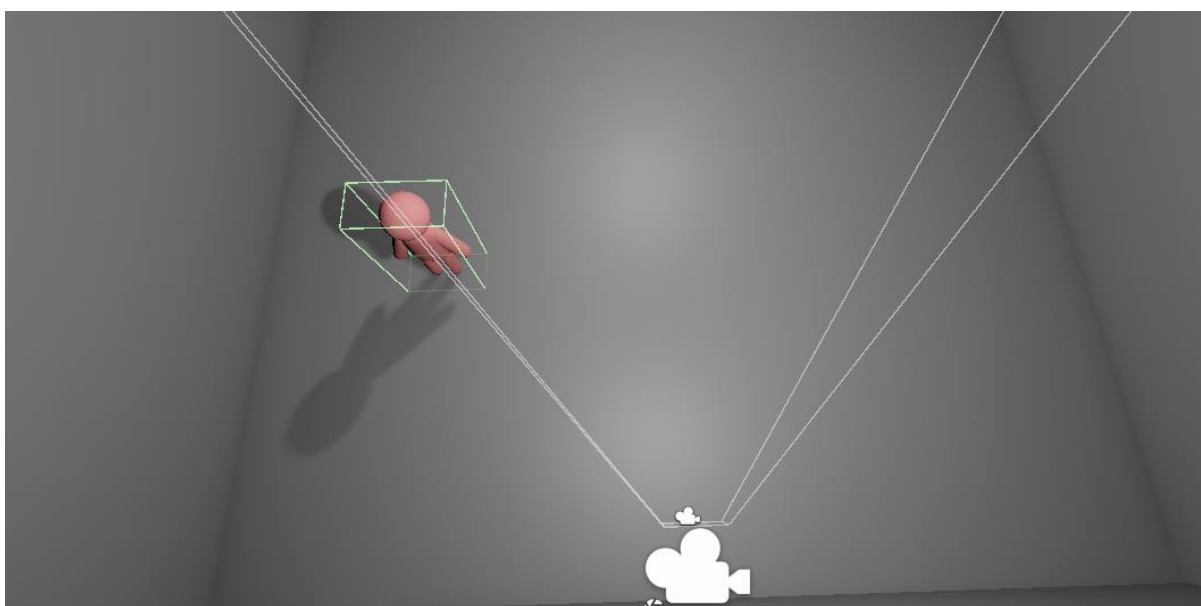


Рис. 1. Пересечение усечённой пирамиды поля зрения и ограничительной рамки цели

2. Угол между вектором, обозначающим направление взгляда персонажа, и вектором между персонажем и целевым объектом должен быть меньше, чем половина заранее заданного угла обзора персонажа.
3. Raycast, выпущенный из координат персонажа в сторону целевого объекта, не должен пересечь по пути какой-либо посторонний объект (рис. 2).

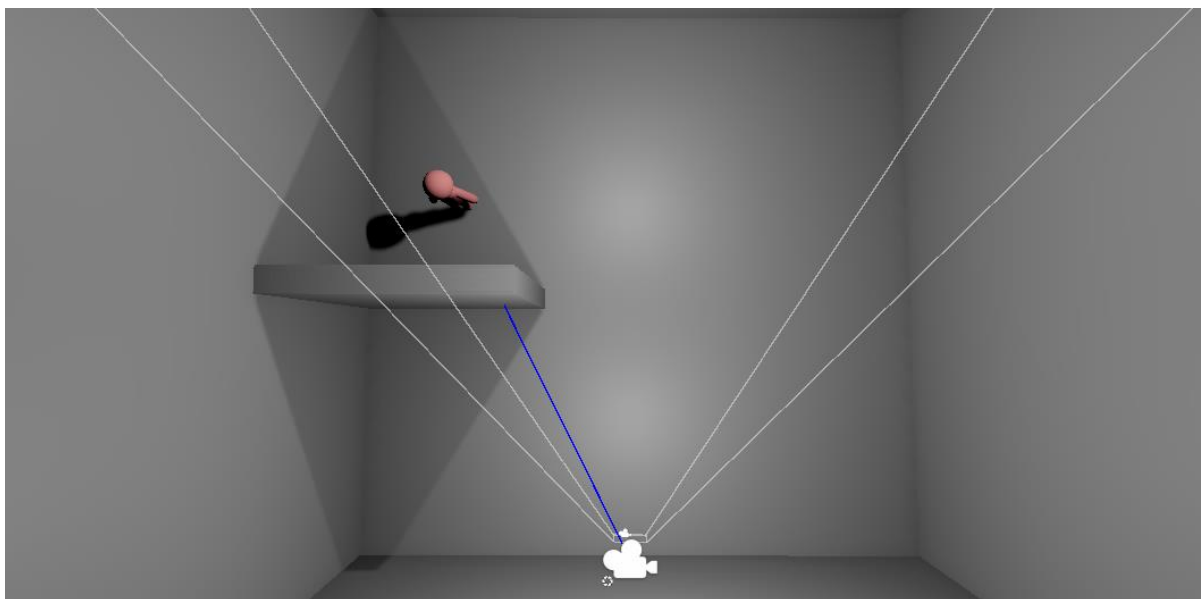


Рис. 2. При трассировке лучей в сторону цели луч сталкивается с препятствием

Метод прост в реализации, однако не учитывает визуальные эффекты, такие как тени или освещение, что снижает его применимость в сложных сценах.

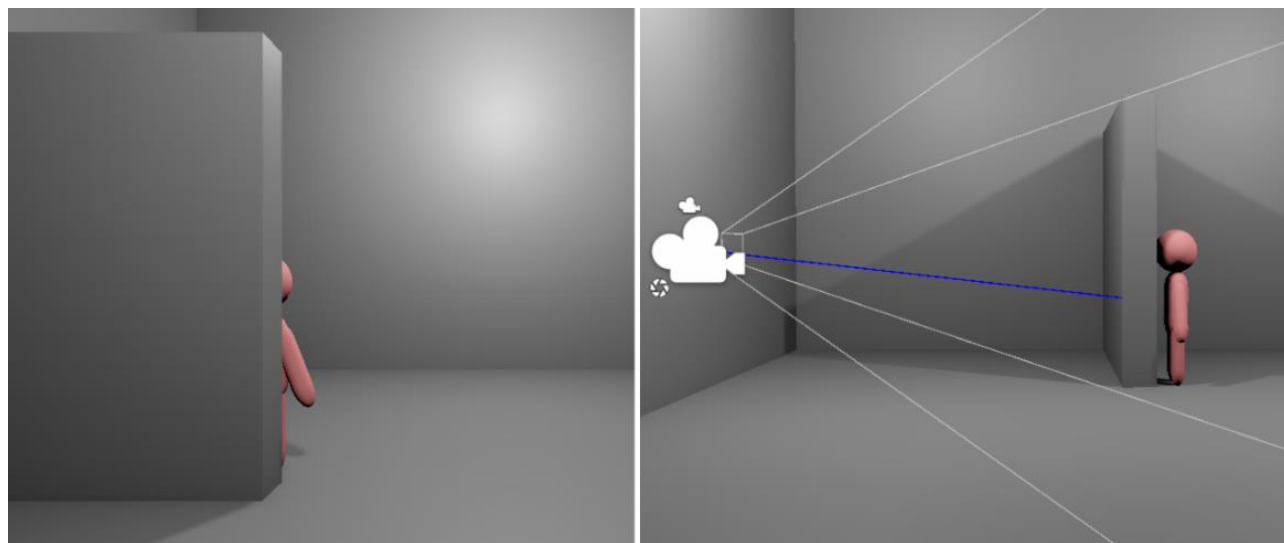


Рис. 3. Raycast заблокирован препятствием несмотря на то, что значительная часть цели видна наблюдателю

Использование одного Raycast может привести к ситуации, когда относительно небольшой объект, например тонкая труба, может заблокировать выпущенный луч несмотря на то, что с точки зрения реального человека персонаж должен видеть небольшую часть целевого объекта (рис. 3).

Другой подход сосредоточен на стелс-механиках, предлагая модель постепенного обнаружения с учетом расстояния и угла обзора [11]. Этот подход повышает реализм в играх с элементами скрытности, но не решает проблему ложных срабатываний, вызванных тенями или отражениями. Авторы отмечают, что мгновенное обнаружение персонажа игрока может показаться нечестным и привести к неудовлетворению игрой, поэтому неигровым персонажам обычно требуется некоторое время, чтобы «заметить» игрока, уже находящегося внутри их поля зрения. По аналогии с реальной жизнью, позиции вблизи неигровых персонажей и прямо перед направлением их взгляда часто связывают с большим риском обнаружения, чем позиции вдали от них и/или на границе их «периферийного зрения».

Плагин NPC Eyes Sight System для Unreal Engine использует точки на скелетных сетках и их проекции на тени, что улучшает обнаружение частично видимых объектов [12]. В такой реализации неигровые персонажи используют для обнаружения других персонажей точки на их скелетных сетках (англ. skeletal mesh). Если искусственный интеллект внутри игры обнаружит в своём поле обзора хотя бы одну такую точку, не прикрытую препятствием, он тут же узнает о местоположении персонажа. Такая реализация значительно уменьшает количество ситуаций, когда небольшое препятствие, лишь частично закрывающее целевой объект, полностью предотвращает его обнаружение персонажами. Однако проецирование множества точек относительно множества возможных источников света – непростая операция, которая может оказать значительное влияние на производительность игры.

Для анализа изображений с камеры NPC в [13] исследовано применение методов машинного зрения, таких как сверточные нейронные сети. Для определения и классификации объектов в поле зрения игрового агента предложено использовать алгоритмы компьютерного зрения. Этот подход обеспечивает высокую точность распознавания, приближая восприятие NPC к человеческому, и позволяет имитировать такие особенности восприятия, как распознавание образов и категоризация объектов. Однако вычислительная сложность этого подхода делает его непригодным в реальном времени без специализированного оборудования, что ограничивает его применимость в обычных игровых проектах.

Для разработки поведения NPC предложен альтернативный подход [14, 15], основанный на применении алгоритмов машинного обучения. Этот метод отходит от жесткого назначения наград и наказаний разработчиком, делегируя эту задачу плагину, который анализирует данные о состоянии агента и автоматически классифицирует поведение на «хорошее» и «плохое» с учетом выбранного целевого параметра. Хотя данный подход направлен на общую систему принятия решений NPC, а неспецифически на визуальное восприятие, его методология динамической настройки поведения в реальной игровой среде может быть адаптирована и для задач визуального восприятия.

Более сложные и реалистичные системы, обеспечивающие правдоподобное поведение NPC, опираются на расширенные подходы: в [16–20] рассмотрены архитектурные решения, позволяющие выбирать тактические позиции и моделировать восприятие в условиях динамической игровой среды. Указанные работы подчёркивают, что ограничение количества тестов видимости в реальном времени может негативно сказаться на многообразии реакций NPC, что оправдывает необходимость разработки более эффективных методов оценки видимости.

Перспективное применение нейронных методов с использованием всенаправленных расстояний (Neural Omnidirectional Distance Fields), представленное в [21], демонстрирует значительное ускорение проверки видимости, что позволяет снизить вычислительную нагрузку по сравнению с традиционными подходами. Это подтверждает актуальность поиска альтернатив классическим raycasting-методам.

Алгоритм, предложенный в настоящей работе, сочетает сравнение текстур, рейкасты и оптимизацию, что отличает его от существующих решений и характеризуется простотой, универсальностью и производительностью.

2. РАЗРАБОТКА АЛГОРИТМА

Для реализации системы зрения игрового агента необходимо определить, видит ли наблюдатель цель в своём поле зрения. Простые методы: проверка углов обзора и расстояния – не учитывают сложные визуальные эффекты, такие как освещение, тени, отражения и маскировка. Поэтому необходим более точный и гибкий подход, который мог бы учитывать визуальные особенности сцены.

2.1. Основной алгоритм

Проблема: традиционные методы определения видимости не учитывают визуальные эффекты, такие как освещение, тени и маскировка, что снижает реализм поведения NPC.

Решение: основной алгоритм системы зрения основан на сравнении изображений, полученных с двух камер, прикреплённых к наблюдателю. Камера 1 рендерит сцену с целью, а Камера 2 — без цели. Обе камеры рендерят ту же сцену, что и основная камера игрока, включая уровень и освещение. Это позволяет учитывать все указанные выше визуальные эффекты (рис. 4).

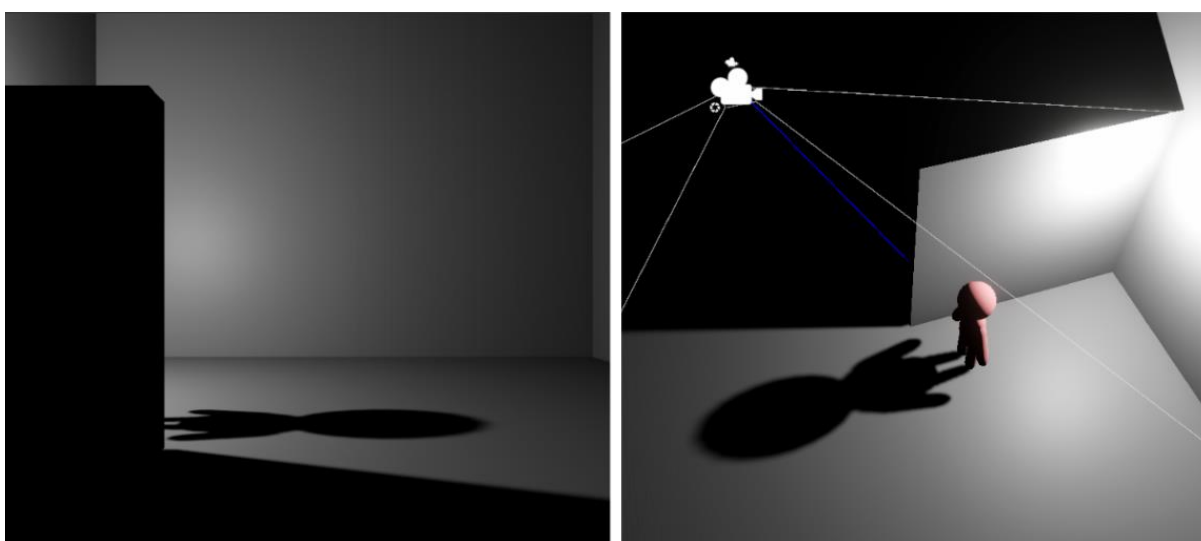


Рис. 4. Наблюдатель видит тень цели, но не саму цель

Шаги алгоритма:

1. Рендеринг сцен в текстуры. Камера 1 включает цель в рендер, а Камера 2 — исключает её.
2. Конвертация текстур с камер в формат Texture2D для дальнейшей обработки.
3. Сравнение пикселей. Вычисление разности пикселей V по каналам RGB и их суммирование для получения степени заметности цели. Сравнение V с пороговым значением $V_{\min}$. Для каждой пары соответствующих пикселей текстур вычисляется их разность по каналам RGB. Эти разности суммируются, чтобы получить степень заметности цели V :

$$V = \frac{\sum_{i=1}^N (|R_1 - R_2| + |G_1 - G_2| + |B_1 - B_2|)}{3N},$$

где R_1, G_1, B_1 – значения каналов RGB для i -го пикселя на 1-ой текстуре, R_2, G_2, B_2 – значения каналов RGB для i -го пикселя на 2-ой текстуре, N – общее количество пикселей.

4. Пороговое значение. Если V превышает пороговое значение $V_{\min}$, считается, что наблюдатель видит цель, т. е. если $V > V_{\min}$, то цель видима.

Преимущества:

- Алгоритм учитывает визуальные особенности сцены (освещение и тени).
- Гибкость: можно настраивать пороговое значение $V_{\min}$ для разных наблюдателей.

Недостатки:

- Высокая вычислительная сложность из-за необходимости рендерить сцену несколько раз и вычислять разность текстур.
- Чувствительность к мелким деталям, что может привести к ложным срабатываниям.

2.2. Маскировка объектов

Проблема: изначальный алгоритм сравнивает отдельные пиксели, что делает его чрезмерно чувствительным к мелким деталям. Например, если цель – это персонаж, одетый в лесной камуфляж на фоне лесной сцены, даже небольшие различия в текстуре могут привести к обнаружению. Это не позволяет объекту «сливаться» с окружающей обстановкой.

Решение: для уменьшения чувствительности к мелким деталям применяется размытие по Гауссу. Размытие снижает резкость изображения, что позволяет объекту маскироваться. Такой подход также является хорошим приближением того, как работает зрительное восприятие человека в реальности [22, 23]. Сложность алгоритма размытия по Гауссу составляет $O(nr^2)$, где n – количество пикселей, а r – радиус круга, по которому усредняется каждый пиксель (другими словами, сила размытия).

Вместо традиционного гауссова фильтра можно применять фильтр Box Blur, который выполняет простое усреднение значений пикселей в квадратном окне.

При нескольких последовательных пропусках Vox Blur позволяет достичь результата, практически неотличимого от размытия по Гауссу, однако имеет линейную вычислительную сложность $O(n)$, что значительно снижает затраты по сравнению с гауссовым фильтром [24].

Шаги решения:

1. Размытие по Гауссу: перед сравнением текстур применяется размытие по Гауссу. Это уменьшает резкость изображения и снижает чувствительность алгоритма к мелким деталям. Радиус определяет степень размытия и обычно устанавливается в диапазоне от 2 до 5 пикселей.

2. Среднеквадратичная ошибка (MSE): вместо простой суммы разностей используется среднеквадратичная ошибка для расчёта V :

$$V = \frac{\sum_{i=1}^N ((R_1 - R_2)^2 + (G_1 - G_2)^2 + (B_1 - B_2)^2)}{3 N},$$

что позволяет уменьшить влияние незначительных изменений в поле зрения и увеличить вес серьёзных различий.

Преимущества:

- Уменьшение чувствительности к мелким деталям.
- Возможность маскировки объектов в сложных условиях (например, камуфляж).
- Увеличение устойчивости к шумам и артефактам.

Недостатки:

- Увеличение вычислительной сложности из-за размытия.
- Необходимость настройки параметров размытия для разных сцен.

2.3. Проверка прямой видимости

Проблема: наблюдатель может реагировать на тени, отражения или малую часть объекта так же, как и на сам объект. Это приводит к ложным срабатываниям, когда наблюдатель «видит» цель, хотя на самом деле видит только её тень или отражение.

Решение: для устранения этой проблемы введена проверка рейкастами. На модели цели равномерно распределяются X точек, которые проверяются на видимость с помощью рейкастов (рис. 5).

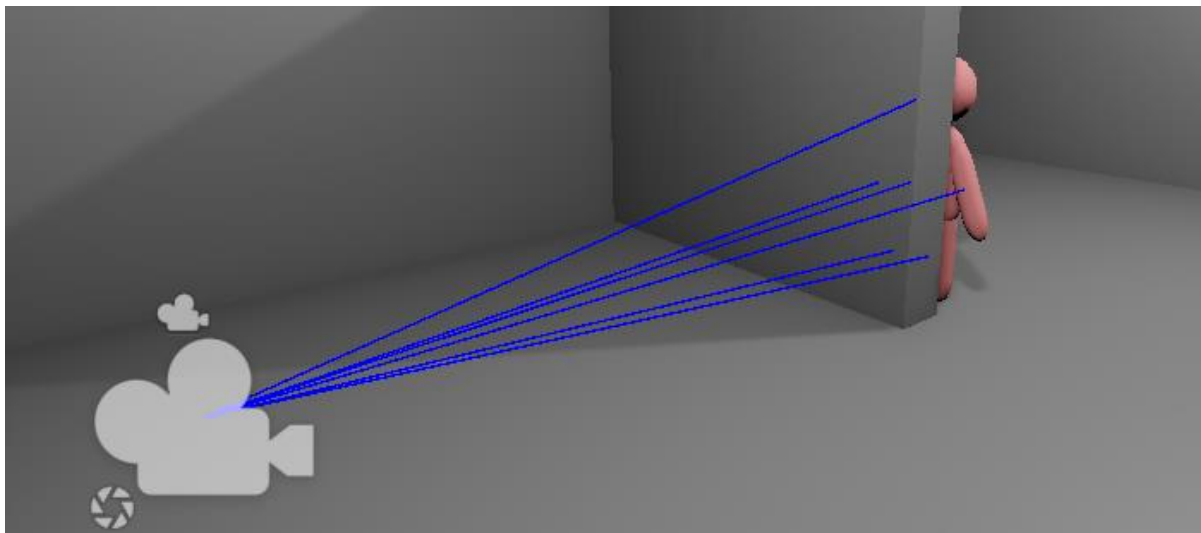


Рис. 5. Пучок из 6 рейкастов, направленных в разные части цели

Если количество X_{pass} точек, прошедших проверку, превышает пороговое значение X_{min} , считается, что наблюдатель видит сам объект. Если $X_{\text{pass}} < X_{\text{min}}$, то наблюдатель видит тень, отражение или малую часть объекта. В этом случае наблюдатель получает информацию о координатах цели и при помощи отдельной логики пытается получить визуальное подтверждение объекта.

Шаги решения:

1. Распределение точек на цели: на модели цели равномерно распределяются X точек, где X обычно принимает значения от 5 до 20 в зависимости от размера и сложности модели.
2. Проверка рейкастами: для каждой точки выполняется рейкаст от наблюдателя до точки. Если луч не пересекает препятствий, точка считается видимой.
3. Определение видимости цели: если количество X_{pass} видимых точек превышает пороговое значение X_{min} , цель считается видимой. В противном случае наблюдатель видит только тень или отражение.

4. Попытка визуального подтверждения: если наблюдатель обнаружил объект, но не имеет прямой видимости, то он попытается получить визуальное подтверждение объекта: например, повернуться в его сторону или же при помощи алгоритма поиска пути дойти до позиции объекта.

Преимущества:

- Уменьшение ложных срабатываний.
- Более реалистичное поведение наблюдателя.
- Возможность настройки порогового значения $X_{\min}$ для разных ситуаций.

Недостатки:

- Увеличение вычислительной сложности из-за необходимости выполнения рейкастов.
- Необходимость настройки количества точек и порогового значения.
- Необходимость равномерной расстановки точек на модели объекта.

2.4. Влияние скорости движения объекта на его заметность

Проблема: в реальности быстро движущиеся объекты заметить легче, чем медленно движущиеся или статичные. Текущий алгоритм не учитывает скорость движения цели, что может привести к нереалистичному поведению наблюдателя.

Решение: для учёта скорости движения цели введён параметр eV – эффективное значение V , который рассчитывается как произведение V на множитель, зависящий от скорости движения цели. Чем выше скорость цели, тем больше множитель, что увеличивает заметность объекта.

Шаги решения:

1. Расчёт скорости цели: определяется скорость движения цели на основе её позиции в предыдущих кадрах.
2. Расчёт множителя: множитель α для V рассчитывается на основе скорости цели. Например, множитель может быть линейно зависимым от скорости.
3. Расчёт V : эффективное значение V рассчитывается как $eV = V\alpha$.
4. Если $eV > V_{\min}$, значит, наблюдатель видит цель.

Преимущества:

- Более реалистичное поведение наблюдателя.
- Динамическая система, учитывающая движение объекта.
- Создает дополнительный геймплейный элемент для стелс-игр, где игрок должен двигаться медленно, чтобы оставаться незамеченным.

Недостатки:

- Необходимость настройки зависимости множителя от скорости.
- Может потребоваться дополнительная балансировка для предотвращения слишком легкого обнаружения быстро движущихся объектов.

2.5. Механика постепенного обнаружения

Проблема: в некоторых случаях обнаружение не должно быть мгновенным. Например, наблюдатель может постепенно «замечать» цель, что добавляет реализма и в случае стелс-игр делает геймплей более честным с точки зрения игрока в ситуации, когда наблюдателями являются противники, а объектом – сам игрок.

Решение: для реализации постепенного обнаружения введён параметр Det – степень обнаружения, который изменяется в зависимости от значения V и скорости обнаружения Det_{speed} .

Если $V > V_{\min}$, то Det увеличивается на $\Delta Det \cdot Det_{\text{speed}}$ каждый кадр. Если $V < V_{\min}$, то Det уменьшается на $\Delta Det \cdot Det_{\text{speed}}$. Когда Det превысит пороговое значение $Det_{\min}$, цель считается обнаруженной.

Шаги решения:

1. Расчёт ΔDet : определяется изменение степени обнаружения на основе значения V .
2. Изменение Det : каждый кадр Det изменяется на $\Delta Det \cdot Det_{speed}$ в зависимости от того, превышает ли V пороговое значение V_{min} .
3. Обнаружение цели: если $Det > Det_{min}$, цель считается обнаруженной.

Преимущества:

- Более реалистичное поведение наблюдателя.
- Возможность настройки скорости обнаружения для разных наблюдателей.
- Улучшенный игровой опыт для стелс-игр, дающий игроку время на реакцию.

Недостатки:

- Необходимость настройки параметров Det_{speed} и Det_{min} .
- Может потребоваться дополнительная логика для реакции наблюдателя на «подозрительный» объект, когда $0 < Det_{min} < Det_{min}$.

2.6. Оптимизация

Проблема: алгоритм имеет высокую вычислительную сложность, что может негативно сказаться на производительности игры, особенно при большом количестве наблюдателей и целей.

Решение: для повышения производительности предложены следующие оптимизации.

1. Использование Unity Job System: алгоритм выполняется в многопоточном режиме с использованием Unity Job System, что позволяет распределить нагрузку на несколько ядер процессора.
2. Уменьшение частоты расчёта V : расчёт V выполняется не в каждом кадре, а с определённым интервалом, что снижает нагрузку на систему.
3. Активация рендера камер: рендер 1 и 2 камер активируется только в тех кадрах, когда необходимо вычислить V , и отключается в остальное время.

4. Исключение заведомо невидимых объектов: алгоритм не применяется для объектов, которые заведомо находятся вне поля зрения наблюдателя (например, слишком далеко или в другой части сцены).

5. Снижение разрешения текстур: для расчёта V используются текстуры с более низким разрешением, чем основная камера игры, что значительно снижает количество пикселей для обработки.

6. Кэширование результатов: результаты рейкастов кэшируются на несколько кадров, если позиции наблюдателя и цели значительно не изменились.

Преимущества:

- Увеличение производительности до 300% по сравнению с наивной реализацией.
- Снижение нагрузки на CPU с 12–15% до 3–5% в среднем.
- Возможность масштабирования для большого количества наблюдателей.

Недостатки:

- Необходимость настройки интервалов расчёта V .
- Усложнение кода из-за многопоточности.
- Потенциальное снижение точности при использовании низкого разрешения текстур.

Помимо описанных оптимизаций, перспективным направлением является применение нейронных методов для аппроксимации проверки видимости. Как показано в [21], использование нейронных Omnidirectional Distance Fields позволяет заменить традиционные рейкасты и ускорить проверку видимости до 9 раз по сравнению с классическим подходом. Этот подход демонстрирует потенциал для дальнейшей оптимизации систем искусственного интеллекта в играх, особенно в сценариях с большим числом наблюдателей.

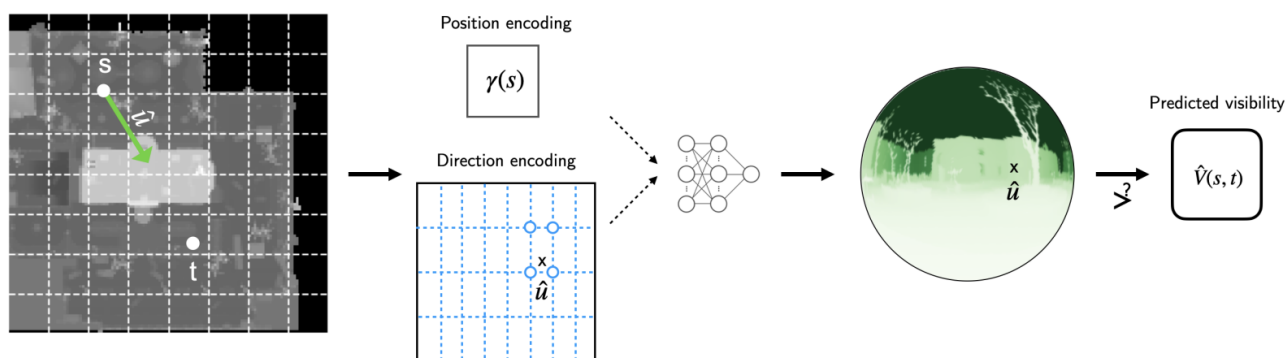


Рис. 6. Подход к аппроксимации видимости между 3D-позициями в статичных игровых сценах с помощью нейронного представления поля всенаправленных расстояний (ODF) [21]

3. ТЕСТИРОВАНИЕ

3.1. Методология тестирования

Для оценки эффективности разработанного алгоритма было проведено комплексное тестирование по нескольким аспектам:

1. Тестирование производительности.
 - Измерение FPS¹ при различном количестве наблюдателей (от 1 до 20).
 - Сравнение с традиционными методами на одинаковых сценах.
 - Профилирование CPU и GPU использования.
2. Тестирование точности обнаружения.
 - Определение процента успешных обнаружений в различных условиях видимости.
 - Оценка частоты ложных срабатываний.
 - Сравнение с ожидаемыми результатами, основанными на восприятии человека.
3. Тестирование в различных игровых условиях.
 - Дневная и ночная сцены с разным освещением.
 - Тестирование маскировки объектов (камуфляж, тени).
 - Сцены с отражающими поверхностями (зеркала, вода).

¹ FPS сокр. от англ. Frame per seconds – количество кадров в секунду.

3.2. Инструменты тестирования

Для проведения тестов были разработаны следующие вспомогательные компоненты.

1. Тепловая карта отличий текстур: визуализирует различие между текстурами с камер 1 и 2, помогая настраивать пороговые значения и улучшать алгоритм.
2. Панель отладки: позволяет во время игры включать/выключать различные компоненты алгоритма (размытие, проверка рейкастами и т. д.), что упрощает тестирование.
3. Система логирования: записывает все случаи обнаружения и параметры наблюдателя и цели для дальнейшего анализа.
4. Сцены для тестирования: созданы специальные сцены, имитирующие различные игровые ситуации, с контролируемыми параметрами освещения и геометрии.
5. Функциональное тестирование: обход технического ограничения стандартного класса Input в Unity путём создания системы автоматизации тестирования, которая позволяет записывать и воспроизводить сценарии взаимодействия пользователя с виртуальной средой [25]. Для тестирования алгоритма визуального восприятия был разработан набор сценариев, моделирующих различные ситуации обнаружения: движение объекта с разной скоростью, перемещение в условиях разной освещенности, попытки маскировки и другие. Применение такого подхода позволило значительно сократить время тестирования за счёт автоматического многократного прогона идентичных тестовых сценариев и сравнения результатов с ожидаемыми, что особенно эффективно для проверки функциональности обнаружения объектов в сложных условиях, когда требуется точная имитация различных входных данных.

3.3. Результаты тестирования

Производительность

Измерения производительности показали, что оптимизированная версия алгоритма добавляет в среднем 3–5% нагрузки на CPU по сравнению с традици-

онным подходом. При этом использование Unity Job System позволило эффективно распределить вычисления между ядрами процессора, что особенно заметно на многоядерных системах. Тестирование на различных устройствах показало, что алгоритм хорошо масштабируется и может быть настроен для разных уровней производительности.

Тестирование точности обнаружения показало значительное улучшение по сравнению с традиционными методами. Разработанный алгоритм значительно лучше работает в сложных условиях видимости и имеет гораздо меньшую частоту ложных срабатываний, когда видны только тени или отражения объекта.

Существенное влияние на эффективность алгоритма оказывает настройка параметров.

1. Влияние радиуса размытия: увеличение радиуса размытия от 2 до 5 пикселей повышало эффективность маскировки на 32%, но снижало точность определения границ объекта.

2. Влияние порогового значения $V_{\min}$: оптимальное значение $V_{\min}$ составило 0.03 для большинства сцен, обеспечив баланс между чувствительностью и устойчивостью к шумам.

3. Влияние количества точек для рейкастов: увеличение числа точек с 5 до 15 повышало точность определения прямой видимости на 27%, но увеличивало нагрузку на CPU на 8%.

4. Влияние $\text{Det}_{\text{speed}}$: значения $\text{Det}_{\text{speed}}$ в диапазоне 0.5–1.0 обеспечивали наиболее реалистичное постепенное обнаружение для стелс-игр.

3.4. Рекомендации по настройке

На основе проведенных тестов были сформулированы следующие рекомендации по настройке параметров алгоритма.

1. Для игр с высоким темпом действия:
 - $V_{\min} = 0.04\text{--}0.05$ (для быстрого обнаружения);
 - $\text{Det}_{\text{speed}} = 1.5\text{--}2.0$ (для быстрой реакции);
 - интервал расчета $V = 0.1\text{--}0.2$ с.;
 - количество точек для рейкастов равно 5–8.

2. Для стелс-игр:

- $V_{\min} = 0.02\text{--}0.03$ (для более тонкого обнаружения);
- $\text{Det}_{\text{speed}} = 0.3\text{--}0.7$ (для постепенного обнаружения);
- интервал расчета $V = 0.2\text{--}0.3$ с.;
- количество точек для рейкастов равно 10–15.

3. Для игр с ограниченными ресурсами (мобильные платформы):

- коэффициент уменьшения разрешения равен 3–4;
- пропуск проверки прямой видимости для отдаленных объектов;
- увеличение интервала расчета V до 0.4–0.5 с.;
- использование более простого алгоритма размытия.

4. РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

Разработанный алгоритм успешно учитывает визуальные эффекты, такие как освещение и маскировка, и обеспечивает гибкость настройки для различных типов NPC. Оптимизация позволяет применять его в реальном времени, однако тестирование выявило необходимость калибровки параметров в зависимости от сцены.

4.1. Сравнение с существующими решениями

Сравнение (табл. 1) с традиционными методами [10] показало, что предложенный подход превосходит их по реализму, но уступает в простоте реализации. В отличие от решений с машинным зрением [13], алгоритм не требует мощного оборудования, что делает его более доступным.

Таблица 1. Сравнение с существующими решениями

Характеристика	Традиционный подход [10]	NPC Eyes Sight System [12]	CV [13]	Разработанный алгоритм
Учет освещения и теней	низкий	средний	высокий	высокий
Реакция на маскировку и камуфляж	низкий	низкий	высокий	средний
Различение объекта и его тени/отражения	низкий	средний	высокий	высокий
Вычислительная сложность	низкая	средняя	очень высокая	средняя
Настраиваемость для разных типов NPC	средняя	средняя	высокая	высокая

4.2. Преимущества разработанного алгоритма

1. Универсальность: алгоритм может быть настроен для различных игровых жанров и условий видимости.
2. Учет визуальных эффектов: система учитывает освещение, тени, отражения и маскировку, что значительно повышает реализм.
3. Оптимизация: благодаря предложенным методам оптимизации, алгоритм может эффективно работать даже на устройствах с ограниченными ресурсами.
4. Постепенное обнаружение: механика постепенного обнаружения делает поведение NPC более реалистичным и улучшает игровой опыт в стелс-играх.
5. Гибкость настройки: широкий набор параметров позволяет тонко настроить систему для каждого типа NPC и игровой ситуации.

4.3. Ограничения и направления дальнейшего развития

Несмотря на значительные преимущества, у разработанного алгоритма есть ряд ограничений:

1. Зависимость от качества рендеринга: Качество работы алгоритма зависит от графического движка и качества рендеринга сцены.

2. Необходимость настройки параметров: Для оптимальной работы в различных сценах требуется тщательная настройка параметров.

3. Ограниченная применимость для динамически изменяющихся условий: при резком изменении освещения или окружения может потребоваться перенастройка параметров.

Направления дальнейшего развития алгоритма включают:

1. Адаптивная настройка параметров: разработка механизма автоматической подстройки параметров в зависимости от условий сцены.

2. Интеграция с другими системами восприятия: комбинирование визуального восприятия со слуховым и другими видами сенсоров для более комплексного поведения NPC.

3. Применение методов машинного обучения: использование упрощенных нейронных сетей для классификации видимых объектов без значительного увеличения вычислительной нагрузки.

ЗАКЛЮЧЕНИЕ

Представлен новый подход к реализации системы визуального восприятия для неигровых персонажей в видеоиграх. Разработанный алгоритм основан на сравнении изображений с двух камер и учитывает сложные визуальные эффекты (освещение, тени и маскировка), что значительно повышает реализм поведения NPC.

Ключевые достижения работы включают:

1. Разработку основного алгоритма сравнения изображений для определения видимости объектов.

2. Реализацию дополнительных механик, таких как учет маскировки, проверка прямой видимости, влияние скорости движения и постепенное обнаружение.

3. Оптимизацию алгоритма для работы в реальном времени с использованием Unity Job System, динамической активации камер и других методов.

4. Всестороннее тестирование алгоритма в различных условиях и формулирование рекомендаций по его настройке.

Результаты тестирования показали, что разработанный алгоритм обеспечивает более реалистичное поведение NPC по сравнению с традиционными методами, особенно в сложных условиях видимости, при этом сохраняя приемлемую производительность.

Предложенный подход может найти применение в различных жанрах видеоигр, особенно в играх с элементами скрытности, где реалистичное визуальное восприятие NPC является критическим фактором для создания увлекательного игрового процесса.

Дальнейшие исследования могут быть направлены на адаптивную настройку параметров алгоритма, интеграцию с другими системами восприятия и применение методов машинного обучения для улучшения классификации видимых объектов.

СПИСОК ЛИТЕРАТУРЫ

1. *Миллингтон Я.* Искусственный интеллект для игр / Я. Миллингтон, Дж. Фанж. СПб.: Питер, 2021. 816 с.
2. *Рабинович З.Л.* Методы и алгоритмы искусственного интеллекта в компьютерных играх: учеб. пособие. М.: Физматлит, 2018. 320 с.
3. *Петров А.В.* Искусственный интеллект в трехмерных играх. Программирование и моделирование поведения персонажей. М.: ДМК Пресс, 2019. 452 с.
4. *Buckland M.* Programming Game AI by Example. Sudbury: Jones & Bartlett Learning, 2022. 522 с.
5. *Ostuni D., Galante E.T.* Towards an AI playing Touhou from pixels: a dataset for real-time semantic segmentation // 2021 IEEE Conference on Games (CoG). 2021. P. 1–5. <https://doi.org/10.1109/CoG52621.2021.9619112>.
6. *Tutum C., AbdulQuddos S., Miikkulainen R.* Generalization of Agent Behavior through Explicit Representation of Context // 2021 IEEE Conference on Games (CoG). 2021. P. 1–7. <https://doi.org/10.1109/CoG52621.2021.9619141>.
7. *Guerrero-Romero C., Perez-Liebana D.* MAP-Elites to Generate a Team of Agents that Elicits Diverse Automated Gameplay // 2021 IEEE Conference on Games (CoG). 2021. P. 1–8. <https://doi.org/10.1109/CoG52621.2021.9619142>.
8. *Glassner A.S. (Ed.).* An introduction to ray tracing. Morgan Kaufmann, 1989.

9. *Panwar H.* The NPC AI of the last of us: a case study // arXiv preprint arXiv:2207.00682. 2022.
10. *Mahmoud I., Jaffal Y., Wloka D.* A Vision Simulation Algorithm for Non-Player Character in Static Scene // University of Kassel, Germany. 2014. P. 6.
11. *Tremblay J., Torres P.A., Verbrugge C.* Measuring Risk in Stealth Games // Foundations of Digital Games. Liberty of the Seas, 2014. P. 8.
12. NPC Eyes Sight System – PRO.
URL: <https://www.fab.com/listings/6b54716a-dd21-414d-b78f-384068de14b7>
13. *Erdelyi C.* Using Computer Vision Techniques to Play an Existing Video Game // California State University San Marcos. 2019. P. 49.
14. *Паренюк Л.Н., Кугуракова В.В.* Разработка плагина поведения NPC для игрового движка Unity // Электронные библиотеки. 2020. Т. 23(5). С. 1044–1057. <http://doi.org/10.26907/1562-5419-2020-23-5-1044-1057>.
15. *Estgren M.* Modelling NPC perception using supervised learning // Uppsala University, Sweden. 2021. 8 p. URL: <https://sciion.se/assets/papers/npc-perception.pdf>
16. *Bourg D.M., Seemann G.* AI for Game Developers. O'Reilly Media, Inc. 2004.
17. *Jack M.* Tactical Position Selection: An Architecture and Query Language. In Game AI Pro 360. CRC Press. 2019. P. 1–24.
18. *McIntosh T.* Human Enemy AI in The Last of Us. In Game AI Pro 360. CRC Press, 2019. P. 13–24.
19. *Welsh R.* Crytek's Target Tracks Perception System. In Game AI Pro: Collected Wisdom of Game AI Professionals. 2013. Vol. 403. 411 p.
20. *Walsh M.* Modeling Perception and Awareness in Tom Clancy's Splinter Cell Blacklist. In Game AI Pro 360. CRC Press. 2019. P. 73–86.
21. *Ying Z., Edwards N., Kutuzov M.* Efficient Visibility Approximation for Game AI using Neural Omnidirectional Distance Fields // Proceedings of the ACM on Computer Graphics and Interactive Techniques. 2024. Vol. 7, No. 1. P. 1–15.
22. Image Comparison Tuned to Human Perception // Computer Science Stack Exchange. 2015.
URL: <https://cs.stackexchange.com/questions/48862/image-comparison-tuned-to-human-perception>
23. *Pramod R.T., Katti H., Arun S.P.* Human peripheral blur is optimal for object

recognition // Vision Research. 2022. Vol. 200. P. 108083.

24. Fastest Gaussian Blur (in Linear Time) // Algorithms and Stuff. 2014.

URL: <https://blog.ivank.net/fastest-gaussian-blur.html>

25. Кузуркова В.В., Бедрин О.А. Система автоматизации функционального тестирования для платформы Unity // Вестник компьютерных и информационных технологий. 2020. Т. 17, № 12. С. 47–52.

<https://doi.org/10.14489/vkit.2020.12.pp.047-052>

DEVELOPMENT OF A VISUAL PERCEPTION SYSTEM FOR GAME AGENTS IN VIDEO GAMES

A. M. Primachenko¹ [0009-0008-6396-2035], **M. R. Khafizov**² [0000-0001-7275-9102]

^{1, 2}*Kazan Federal University, Kazan, 420008, Russia*

¹primachenko.artem@mail.ru, ²murkorp@gmail.com

Abstract

The developed algorithm of the visual perception system for game agents, implemented in the Unity game engine, is presented. The proposed method is based on the comparison of images from two cameras, taking into account complex visual effects (lighting, shadows, camouflage), and supplemented with line-of-sight verification, taking into account the speed of the object, and the mechanics of gradual detection. Testing of the system has shown a significant increase in realistic detection compared to traditional methods, while maintaining performance within a small additional load on the processor. The algorithm was optimized using Unity Job System and dynamic camera activation. The scientific literature on similar solutions has also been analyzed and their strengths and weaknesses have been identified. The results can be applied in video game development to create realistic behavior of non-player characters, especially in games with stealth elements.

Keywords: *video games, artificial intelligence, perception system, NPC, non-player characters, game agents, stealth mechanics, Unity, rendering, computer vision, optimization, game design.*

REFERENCES

1. *Millington I., Funge J.* Artificial Intelligence for Games. Saint Petersburg: Piter. 2021. 816 p. (in Russian).
2. *Rabinovich Z.L.* Methods and Algorithms of Artificial Intelligence in Computer Games: A Tutorial. Moscow: Fizmatlit. 2018. 320 p. (in Russian).
3. *Petrov A.V.* Artificial Intelligence in 3D Games: Programming and Modelling of Character Behavior. Moscow: DMK Press. 2019. 452 p. (in Russian).
4. *Buckland M.* Programming Game AI by Example. Sudbury: Jones & Bartlett Learning, 2022. 522 p.
5. *Ostuni D., Galante E.T.* Towards an AI playing Touhou from pixels: a dataset for real-time semantic segmentation // 2021 IEEE Conference on Games (CoG). 2021. P. 1–5. <https://doi.org/10.1109/CoG52621.2021.9619112>.
6. *Tutum C., AbdulQuddos S., Miikkulainen R.* Generalization of Agent Behavior through Explicit Representation of Context // 2021 IEEE Conference on Games (CoG). 2021. P. 1–7. <https://doi.org/10.1109/CoG52621.2021.9619141>.
7. *Guerrero-Romero C., Perez-Liebana D.* MAP-Elites to Generate a Team of Agents that Elicits Diverse Automated Gameplay // 2021 IEEE Conference on Games (CoG). 2021. P. 1–8. <https://doi.org/10.1109/CoG52621.2021.9619142>.
8. *Glassner A.S. (Ed.).* An introduction to ray tracing. Morgan Kaufmann, 1989.
9. *Panwar H.* The NPC AI of the last of us: a case study. arXiv preprint arXiv:2207.00682. 2022.
10. *Mahmoud I., Jaffal Y., Wloka D.* A Vision Simulation Algorithm for Non-Player Character in Static Scene // University of Kassel, Germany. Kassel, 2014. P. 6.
11. *Tremblay J., Torres P.A., Verbrugge C.* Measuring Risk in Stealth Games // Foundations of Digital Games. Liberty of the Seas, 2014. P. 8.
12. NPC Eyes Sight System PRO.
URL: <https://www.fab.com/listings/6b54716a-dd21-414d-b78f-384068de14b7>
13. *Erdelyi C.* Using Computer Vision Techniques to Play an Existing Video Game // California State University San Marcos. 2019. P. 49.
14. *Parenyuk L.N., Kugurakova V.V.* NPC behavior plugin development for game engine Unity // Russian Digital Libraries. 2020. Vol. 23 (5). P. 1044–1057. <https://doi.org/10.26907/1562-5419-2020-23-5-1044-1057> (In Russian).

15. *Estgren M.* Modelling NPC perception using supervised learning // Uppsala University, Sweden. 2021. 8 p.
URL: <https://sciion.se/assets/papers/npc-perception.pdf>
16. *Bourg D.M., Seemann G.* AI for Game Developers. O'Reilly Media, Inc. 2004.
17. *Jack M.* Tactical Position Selection: An Architecture and Query Language. In Game AI Pro 360. CRC Press. 2019. P. 1–24.
18. *McIntosh T.* Human Enemy AI in The Last of Us. In Game AI Pro 360. CRC Press, 2019. P. 13–24.
19. *Welsh R.* Crytek's Target Tracks Perception System. In Game AI Pro: Collected Wisdom of Game AI Professionals. 2013. Vol. 403. 411 p.
20. *Walsh M.* Modeling Perception and Awareness in Tom Clancy's Splinter Cell Blacklist. In Game AI Pro 360. CRC Press. 2019. P. 73–86.
21. *Ying Z., Edwards N., Kutuzov M.* Efficient Visibility Approximation for Game AI using Neural Omnidirectional Distance Fields // Proceedings of the ACM on Computer Graphics and Interactive Techniques. 2024. Vol. 7, No. 1. P. 1–15.
22. Image Comparison Tuned to Human Perception // Computer Science Stack Exchange. 2015.
URL: <https://cs.stackexchange.com/questions/48862/image-comparison-tuned-to-human-perception>
23. *Pramod R.T., Katti H., Arun S.P.* Human peripheral blur is optimal for object recognition // Vision Research. 2022. Vol. 200. P. 108083.
24. Fastest Gaussian Blur (in Linear Time) // Algorithms and Stuff. 2014.
URL: <https://blog.ivank.net/fastest-gaussian-blur.html>
25. *Kugurakova V.V., Bedrin O.A.* Functional test automation system for unity applications // Vestn. Komp'yut. Inform. Tekhnol. 2020. Vol. 17, No. 12. P. 47–52. <https://doi.org/10.14489/vkit.2020.12.pp.047-052> (In Russian).

СВЕДЕНИЯ ОБ АВТОРАХ



ПРИМАЧЕНКО Артём Михайлович – магистрант Института информационных технологий и интеллектуальных систем Казанского федерального университета. Направление исследований: визуальное восприятие игровых агентов в видеоиграх.

Artyom Mikhailovich PRIMACHENKO – Master’s student at the Institute of Information Technologies and Intelligent Systems of Kazan Federal University. Research area: visual perception of game agents in video games.

email: primachenko.artem@mail.ru

ORCID: 0009-0008-6396-2035



ХАФИЗОВ Мурад Рустэмович – старший преподаватель Института ИТИС КФУ. Сфера научных интересов: разработка видеоигр, виртуальная реальность, синтетические данные.

Murad Rustemovich KHAFIZOV – senior lecturer at the Institute of Information Technologies and Intelligent Systems, Kazan Federal University. Research interests: game development, virtual reality, synth data.

e-mail: murkorp@gmail.com

ORCID: 0000-0001-7275-9102

Материал поступил в редакцию 2 февраля 2025 года

THE EDUCATIONAL VIDEO GAME FOR THE PROMOTION OF TRADITIONAL CUBAN MUSIC AMONG CHILDREN AND TEENAGERS

A. de J. Parada Figueroa¹ [0009-0004-7332-9883], O. Correa Madrigal² [0000-0002-1135-685X]

^{1, 2}*Center for Interactive Technologies,
University of Computer Sciences (UCI), Havana, Cuba*

¹abrahamweer@gmail.com, ²ocorrea@uci.cu

Abstract

This article presents the design and development of the educational video game, conceived as a tool for the promotion and preservation of traditional Cuban music of the Son genre among children and adolescents. The proposal combines cultural content with contemporary game mechanics, generating an immersive and didactic experience. Its pedagogical foundations, playful structure, and potential impact on the formation of cultural awareness from an early age are described.

Keywords: *video game, educational video game, gamification, culture, traditional Cuban music, Cuba, interactive technologies, playful teaching.*

INTRODUCTION

Traditional Cuban music constitutes an essential component of the nation's intangible heritage, acting as a vehicle for collective memory, cultural identity, and social cohesion. Among its most representative expressions is the They are Cuban, a musical genre that synthesizes African and Hispanic influences and whose evolution has accompanied the country's historical and sociopolitical processes. However, in the contemporary context, marked by increasing media globalization and the predominance of foreign cultural products, there is a progressive disconnect between new generations and these indigenous musical expressions [1].

This generational-cultural gap is accentuated by the scarcity of educational resources that, using an updated pedagogical approach, manage to convey this content in a meaningful, playful way that is in tune with the communication codes of today's youth. The dissociation between traditional teaching formats and contemporary forms of cultural consumption poses an urgent challenge for the active preservation of heritage [2].

In response to this problem, the project *The Route of the Son* developed in collaboration with the Center for Interactive Technologies of the University of Informatics Sciences (UCI), proposes the creation of a recreational-educational platform aimed at promoting Cuban genre of music through the interactive language of video games. This initiative stems from the recognition of video games not only as a means of entertainment, but also as a cultural mediation tool and pedagogical device, capable of activating processes of symbolic appropriation, meaningful learning and identity construction.

The project is part of an emerging movement in Latin America that seeks to articulate technology, education, and local culture through innovative, inclusive, and participatory approaches. In this sense, this research not only describes the video game design process but also analyzes its theoretical foundations, methodological strategies, and potential impact in educational and community contexts. It thus proposes a contribution to the interdisciplinary field of video game studies – educational video games with a cultural focus, with particular emphasis on the socio-technological dimension of intangible heritage [3].

The Route of the Son its main objective is to stimulate interest in traditional Cuban music among children and adolescents through an interactive adventure. Players symbolically travel across the country on the so-called “Son Island,” discovering the history, instruments, performers, and songs that shaped this iconic genre.

DIDACTIC DESIGN AND GAME MECHANICS

The video game's educational design is based on a modular structure composed of a set of interrelated minigames, each with specific learning objectives and distinct mechanics that address diverse learning styles (visual, auditory, kinesthetic). This playful architecture responds to a logic of progression in which the user, by overcoming challenges, accesses new content and levels of increasing complexity, thus promoting meaningful learning and sustained motivation [4].

Below are the main game mechanics, as well as their function within the video game's overall educational system:

1. Trivia (questions and answers)

Type of mechanics: Cognitive / Multiple Choice

Pedagogical objective: Evaluate and reinforce acquired knowledge about the history, exponents, and sociocultural contexts of Son.

Logical structure:

- Selection interface with several questions per level.
- Each question has 3 or 4 answer options, of which only one is correct.
- Immediate feedback with brief textual feedback (reinforcement or clarification).
- Scoring and time system to encourage efficient performance.

Educational justification: Stimulates declarative memory and critical thinking through the recognition and retrieval of contextualized information.

2. Runner musical

Type of mechanics: Skill / Reaction time / Motor coordination

Pedagogical objective: Reinforce active listening through interaction with musical elements and cultural.

Logical structure:

- The character automatically moves through an environment that represents sound and visual landscapes of Cuba.
- Obstacles appear synchronized with the rhythm of a Son piece, which the player must avoid or use strategically.
- Collecting themed objects (instruments, records, cultural icons) that award points or additional information.
- Educational justification: It increases musical sensitivity, fosters sustained attention, and strengthens the emotional connection to cultural content through sensory immersion.

3. Memory (card association)

Type of mechanics: Visual and auditory memory / Semantic association

Pedagogical objective: Stimulate working memory by associating elements related to musical heritage.

Logical structure:

- Board with hidden cards that the player must turn over in turns.
- When pairs are found, a short animation or explanatory musical fragment is activated.
- Pairs can combine images (instruments, characters) with sounds, texts either dates.

Educational justification: It reinforces the recognition and conceptual association between different representations of Sound, integrating multisensory stimuli.

4. Piano Tiles (adapted musical version)

Type of mechanics: Coordination / Rhythm / Tactile or keyboard interaction

Pedagogical objective: Facilitate melodic recognition and the characteristic rhythm of Son through active performance.

Logical structure:

- Descending panel where “keys” (notes or measures) fall to the rhythm of a traditional song.
- The player must press them at the exact moment to keep the melody flowing.
- Free practice or scoring execution modalities, with feedback on accuracy rhythmic.

Educational justification: It involves the body in the musical learning process, connecting perception, action and implicit knowledge.

INNOVATION AND CULTURAL APPLICATION

The video game is conceived as a cultural mediation tool that promotes immersive and playful learning processes, articulating heritage content with interactive dynamics aimed at identity building. Through visual and audio elements characteristic of the Cuban imagination, the proposal fosters not only knowledge retention but also the symbolic appropriation of heritage from an early age. This methodological strategy, based on gamification, allows for the exploration of the relationship between culture, education, and technology, while contributing to the development of a sense of belonging in different population segments. The thematic selection, centered on indige-

nous cultural expressions, strengthens the connection between subjects and their sociocultural environment, expanding the scope of learning beyond the school setting [5].

SOCIO-CULTURAL APPROACH AND POPULATION PROJECTION

The video game is conceived as a cultural mediation tool that promotes immersive and playful learning processes, articulating heritage content with interactive dynamics aimed at building identity in diverse social contexts. Far from being limited to a one-way transmission of knowledge, the gaming experience stimulates the subject's active participation, facilitating processes of meaning-making where culture is experienced, recreated, and reinterpreted.

Through a careful integration of visual, audio, and narrative elements representative of Cuban culture—such as traditional musical rhythms, iconic urban and rural settings, and archetypal characters—a symbolic environment is constructed that fosters both knowledge retention and the critical appropriation of cultural heritage. The methodological proposal, based on gamification, transforms learning into a dynamic, motivating experience deeply connected to users' sociocultural reality [6].

From a population perspective, video games are not only targeted at a specific age group, but also allow for the exploration of identity processes in different social groups, including children, adolescents, and young adults. By fostering the recognition and appreciation of one's own cultural expressions, they contribute to strengthening community ties, promote a sense of belonging, and stimulate reflection on the role of culture in building citizenship.

In this sense, innovation lies not only in the incorporation of digital technologies, but in their strategic use to activate teaching and learning processes based on local cultural references. This is an expanded pedagogical experience, where play and symbolism are articulated within an educational project with a territorial impact and the potential for replication in other sociocultural contexts [7].

GENERAL STRUCTURE OF THE GAME SYSTEM

The minigames are integrated into a narrative progression system where the player travels through different regions of Cuba, each associated with specific cultural content. Access to new regions depends on completing objectives within the

minigames, allowing for constant feedback between gameplay progress and knowledge acquisition.

Additionally, it incorporates an achievement system, an educational log (where acquired knowledge and unlocked curiosities are stored), and an accessible interface that respects universal design principles to ensure inclusion and ease of use.

STYLE, GAME DESIGN, AND IMMERSIVE EXPERIENCE

The video game's visual and audio style responds to a logic of cultural coherence, seeking to generate an immersive atmosphere that speaks directly to the roots of Son as an artistic expression and an identity symbol of Cuban culture. From a user-centered design perspective, the game's overall aesthetic has been carefully crafted to evoke the warmth, rhythm, and vitality inherent to Son's symbolic universe, utilizing a warm color palette, traditional settings, and visual elements reminiscent of neighborhoods, plazas, traditional musical instruments, and traditional clothing.

This approach allows to articulate a immersive and traditional gameplay, in which the player not only “controls” an avatar, but rather subjectively projects themselves into a playful environment that feels familiar, emotional, and evocative. This immersion is enhanced by an original or adapted soundtrack, which reproduces representative styles of Son, integrating harmonic progressions, rhythmic patterns, and native timbres such as the tres, marímbula, and claves.

Pedagogical approach to design

The game system incorporates contemporary pedagogical techniques based on the gamification (gamification) and the experiential learning, prioritizing the player's active participation in knowledge construction. Interaction is not limited to responding to stimuli or overcoming challenges, but rather promotes exploration, discovery, association, and symbolic experimentation, establishing bridges between cultural information and emotional experience.

Every component of the design – levels, mechanics, rewards, progression – aligns with principles of meaningful learning, which allows the content to not be perceived as external or obligatory, but as an integral part of an experience that combines entertainment with the appropriation of knowledge.

Emotional and cultural integration of the player

One of the key principles of design is the high integration with the target audience, conceived not as a passive recipient, but as a cultural actor with agency. The goal is for the player to recognize themselves in the content, be moved by the music, identify with the landscapes and characters, and actively participate in the rescue and revaluation of heritage. In this sense, the game not only teaches about Son, but also contributes to revitalize collective memory, to strengthen the sense of belonging and foster pride in the shared cultural heritage [8, 9].

The design, therefore, aims for an expanded transmedia experience: what the player experiences within the virtual environment can be transferred to practices, knowledge, and values outside the game, multiplying its educational and sociocultural impact.

PROJECTION AND INNOVATION AT A REGIONAL KEY

This video game is positioned as a pioneering proposal within the emerging landscape of Cuban and Latin American educational video games, aimed at rescuing, preserving, and disseminating cultural values from playful, inclusive, and technologically advanced perspectives. Its local-centric approach — both in terms of content and aesthetic, musical, and symbolic references—constitutes a deliberate act of creative resistance to the hegemony of globalized narratives and a firm commitment to cultural sovereignty in the digital realm.

In this context, innovation is not limited to the technical or graphic, but resides in its capacity to propose an alternative model of education and entertainment that engages with the realities, knowledge, and emotions of the communities it addresses. This model integrates affective, cognitive, and social dimensions and aligns with the tenets of critical education, situated learning, and cultural pedagogy.

From a medium and long term perspective, the game establishes a methodological and conceptual precedent for future developments focused on other cultural expressions endemic to the region. In doing so, it opens a path for the Construction of a Latin American ecosystem of educational video games, in which cultural diversity, digital inclusion and community participation function as the articulating axes of design, implementation and evaluation.

Furthermore, its replicable and adaptable potential makes it a strategic benchmark for educational institutions, cultural centers, and developers interested in creating products that combine pedagogical value, sociocultural relevance, and emotional impact.

In short, the video game not only represents an innovation in the field of educational technology, but also acts as a intergenerational transmission vehicle, a platform for symbolic empowerment and a catalyst for new ways of imagining learning from our own, our local, and our meaningful perspective.

CONCLUSIONS

The development of a recreational-educational platform focused on Cuban Son, as *The Route of the Sound*, demonstrates the potential of video games as cultural mediators to revitalize intangible heritage by generating experiences that foster symbolic appropriation and identity recognition in children and young people. This proposal articulates gamification techniques with a pedagogical design based on meaningful learning and multisensory stimulation, integrating musical and cultural knowledge in an engaging way and adapted to contemporary communication codes, which favors lasting learning that is emotionally linked to the player. Beyond its technological dimension, the video game emerges as a sociocultural intervention strategy with territorial impact, adaptable to school, community, and museum contexts thanks to its modular structure and inclusive approach, reinforcing the link between culture, citizenship, and technology. Likewise, the video game *The Route of the Son* It is positioned as a replicable benchmark in Latin America for preserving endemic cultural expressions through interactive technologies, promoting a regional ecosystem of educational video games with its own identity. Finally, this initiative contributes interdisciplinarily to the field of educational technology by integrating critical pedagogy, digital design, and cultural studies, offering conceptual and practical foundations for future research exploring the transformative use of ICTs in heritage education.

REFERENCES

1. *Gonzalez Hernandez A.* Cuban Son: History and Cultural Legacy. 2018. 320 p.
2. *Martinez L.* Gamification and Cultural Heritage: Strategies for Interactive Education // RIVERA. Digital Technologies in Cultural Preservation. TecnoCultura Editions. 2021. P. 89–114.
3. *Sanchez A.* Video Games as Tools for the Transmission of Intangible Heritage: The Case of Cuban Son // Ibero-American Journal of Digital Education. 2023. Vol. 15. No. 3. P. 45–67.
4. *Dondlinger M. J.* Educational video game design: A review of the literature // Journal of applied educational technology. 2007. Vol. 4. No. 1. P. 21–31. URL: https://ocw.metu.edu.tr/pluginfile.php/2390/mod_resource/content/0/ceit706/week8/EducationalGameDesign.pdf
5. *Ma Y., Wang C.* Empowering music education with technology: a bibliometric perspective // Humanities and Social Sciences Communications. 2025. Vol. 12. No. 1. P. 1-14. URL: <https://doi.org/10.1057/s41599-025-04616-2>
6. *Fernandez P.* Cuban Son in the Digital Age: Strategies for its Preservation // Culture and Society. 2021. No. 12.
7. *Benge J.F., Scullin M.K.* A meta-analysis of technology use and cognitive aging. Nat Hum Behav (2025). <https://doi.org/10.1038/s41562-025-02159-9>
8. Strategies for the Promotion of Traditional Music among Young People // Cuba's Ministry of Culture: Cultural Editions. 2019. 150 p.
9. *Diaz R.* Cuban Son: Rhythm and Identity // Caribbean Productions. 2020.

ОБУЧАЮЩАЯ ВИДЕОИГРА ДЛЯ ПОПУЛЯРИЗАЦИИ ТРАДИЦИОННОЙ КУБИНСКОЙ МУЗЫКИ СРЕДИ ДЕТЕЙ И ПОДРОСТКОВ

А. де Х. Парада Фигероа¹ [0009-0004-7332-9883], О. Корреа Мадригал² [0000-0002-1135-685X]

^{1, 2}Центр интерактивных технологий,
Университет компьютерных наук (UCI), Гавана, Куба

¹abrahamweer@gmail.com, ²ocorrea@uci.cu

Аннотация

Представлены дизайн и разработка образовательной видеоигры, задуманной как инструмент для продвижения и сохранения традиционной кубинской музыки среди детей и подростков. В игре сочетаются культурный контент и современные игровые механики, создавая иммерсивный и дидактический опыт. Описаны его педагогические основы, игровая структура и потенциальное влияние на формирование культурного сознания с раннего возраста.

Ключевые слова: видеоигры, образовательная видеоигра, геймификация, традиционная музыка, культура, традиционная Кубинская музыка, Куба, интерактивные технологии, игровое обучение.

СПИСОК ЛИТЕРАТУРЫ

1. Gonzalez Hernandez A. Cuban Son: History and Cultural Legacy. 2018. 320 p.
2. Martinez L. Gamification and Cultural Heritage: Strategies for Interactive Education // RIVERA. Digital Technologies in Cultural Preservation. TecnoCultura Editions. 2021. P. 89–114.
3. Sanchez A. Video Games as Tools for the Transmission of Intangible Heritage: The Case of Cuban Son // Ibero-American Journal of Digital Education. 2023. Vol. 15. No. 3. P. 45–67.
4. Dondlinger M. J. Educational video game design: A review of the literature // Journal of applied educational technology. 2007. Vol. 4. No. 1. P. 21–31. URL: https://ocw.metu.edu.tr/pluginfile.php/2390/mod_resource/content/0/ceit706/week8/EducationalGameDesign.pdf

5. *Ma Y., Wang C.* Empowering music education with technology: a bibliometric perspective // *Humanities and Social Sciences Communications*. 2025. Vol. 12. No. 1. P. 1-14. URL: <https://doi.org/10.1057/s41599-025-04616-2>
6. *Fernandez P.* Cuban Son in the Digital Age: Strategies for its Preservation // *Culture and Society*. 2021. No. 12.
7. *Benge J.F., Scullin M.K.* A meta-analysis of technology use and cognitive aging. *Nat Hum Behav* (2025). <https://doi.org/10.1038/s41562-025-02159-9>
8. Strategies for the Promotion of Traditional Music among Young People // Cuba's Ministry of Culture: Cultural Editions. 2019. 150 p.
9. *Díaz R.* Cuban Son: Rhythm and Identity // Caribbean Productions. 2020.

СВЕДЕНИЯ ОБ АВТОРАХ



Абрахам де Хесус ПАРАДА ФИГУЭРОА — Доцент Университета информатики (Куба), руководитель проекта «La Ruta del Son». Сфера научных интересов — образовательные видеоигры.

Abraham de Jesús PARADA FIGUEROA — Assistant Professor of University of Informatics Sciences (Cuba), Head of “Route of Son” project. His area of research interest is educational video games.

email: abrahamweer@gmail.com

ORCID: 0009-0004-7332-9883



КОРРЕА МАДРИГАЛ Омар — профессор, директор Центра интерактивных технологий Университета информатики (Куба). Также приглашенный профессор кафедры индустрии разработки видеоигр Института информационных технологий и интеллектуальных систем Казанского федерального университета. Сфера научных интересов — различные аспекты разработки компьютерных игр и VR-тренажеров.

Omar CORREA MADRIGAL — Dr., Full Professor, Director of the Center for Interactive Technologies at the University of Informatics (Cuba). He is also a visiting professor of the Department of Video Game Development Industry at the Institute of Information Technologies and Intelligent Systems of Kazan Federal University. His research interests include various aspects of computer games and VR simulators development.

email: ocorrea@uci.cu

ORCID: 0000-0002-1135-685X

Материал поступил в редакцию 25 марта 2025 года

ИССЛЕДОВАНИЕ ИГРОВОГО ОПЫТА И ФОРМИРОВАНИЕ КОГНИТИВНОЙ МОДЕЛИ ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЕЙ С НОВОЙ ВИДЕОИГРОЙ

М. В. Шматко¹ [0000-0002-7255-8885], А. А. Кутузов² [0000-0004-6652-2968],
Л. Р. Пономарев³ [0009-0009-9600-7824]

^{1, 2, 3}Омский государственный технический университет, г. Омск, 644050,
Россия

¹marin298@gmail.com, ²kutuzovschool@mail.ru, ³oil294semrom@gmail.com

Аннотация

Статья посвящена исследованию игрового опыта пользователей и формированию когнитивной модели их взаимодействия с новой видеоигрой «Кошечские проделки», находящейся на ранней стадии разработки. Предложенная авторами методология UX-исследования, а также полученные результаты демонстрируют, как качественные данные о восприятии, эмоциях, чувствах, моделях поведения и достижениях игроков помогают оптимизировать процесс создания видеоигр, сделать их не только развлекательными, но и психологически эффективными. Результаты работы имеют практическую ценность для инди-разработчиков, которые, обладая ограниченными ресурсами, стремятся к созданию качественных игровых продуктов.

Ключевые слова: видеоигра, гейм-дизайн, игровой опыт, опыт пользователей, когнитивное взаимодействие, когнитивная модель, UX-исследование.

ВВЕДЕНИЕ

Современный рынок видеоигр характеризуется высокой конкуренцией и растущими ожиданиями со стороны пользователей, что делает критически важным глубокое понимание игрового опыта уже на этапе создания концепции нового продукта. Особую актуальность это приобретает для инди-разработчиков,

чи ресурсы ограничены, а риски коммерческого провала значительно выше по сравнению с крупными студиями.

Важность исследования игрового опыта и формирования когнитивной модели взаимодействия пользователей с видеоигрой обоснована ростом спроса на «осознанный» гейм-дизайн, а также необходимостью культурной адаптации игрового контента. В связи с этим интеграция исследований с разработкой видеоигр становится не просто рекомендацией, а обязательным требованием. Только такой подход может обеспечить коммерческую жизнеспособность игровому проекту и глубину вовлечения пользователей в игровой процесс.

Цель настоящей работы – получить качественные данные об игровом опыте потенциальных пользователей новой видеоигры «Кошечьи проделки» и применить их для формирования когнитивной модели, обеспечивающей глубокую степень вовлеченности пользователей в игровой процесс.

ИГРОВОЙ ОПЫТ И ФРЕЙМВОРКИ ДЛЯ ЕГО ИССЛЕДОВАНИЯ

Понятие «игровой опыт», или «опыт игроков», (англ. game user experience – GUX) является производным понятия «пользовательский опыт» (англ. user experience – UX) и определяется как результат когнитивного взаимодействия игрока с различными игровыми продуктами. По отношению к видеоиграм оно употребляется в контексте взаимодействий, которые осуществляются в системе «человек – машина» («органы чувств – мозг – компьютер») и напрямую влияют на когнитивные функции игрока, трансформируя модели его поведения не только в игре, но и в обычной жизни [1–3].

Так, например, Коллин Маклин и Джон Шарп в своем труде «Игры: дизайн и игровой опыт. Все об итеративной разработке игр» говорят о трансформации игрового опыта в процессе взаимодействия игрока с игрой при погружении в ее события. Саму игру при этом авторы называют «системой-произведением, созданным для выражения, передачи и создания игрового опыта» [4].

Роберт Зубек, автор книги «Элементы гейм-дизайна. Как создавать игры, от которых невозможно оторваться», характеризует опыт совокупностью субъективных впечатлений игрока от «взаимодействия с различными игровыми объектами, правилами, возможно, другими участниками игры» [5].

Исходя из сказанного, важно понять, посредством каких когнитивных функций формируется игровой опыт. Эта информация позволит разработчику видеоигры выбрать эффективные средства воздействия на пользователей и сформировать у них такую модель игрового поведения, которая обеспечит высокую степень вовлеченности в игровой процесс.

Представим сначала уровни когнитивного взаимодействия игрока с видеоигрой (от примитивных к сложным):

- физиологические реакции (прежде всего зрительные и слуховые ощущения);
- восприятие виртуального игрового мира и его наполнения (прежде всего пространственное и художественное);
- эмоции на различные компоненты игры (основа для ее оценки);
- чувства (основа для формирования игровых предпочтений);
- убеждения (основа для определения игровых стратегий);
- модели игрового поведения (основа для выбора игровых тактик);
- достижения в игре (основа для экстраполяции игрового опыта с одной игры на другую) [6, 7].

На каждом из этих уровней игрок производит более сложную обработку информации относительно всех компонентов видеоигры, а его игровой опыт расширяется и трансформируется в результате взаимодействия с новыми видеоиграми [7, 8].

В связи с этим возникает вопрос, можно ли проектировать игровой опыт пользователей? Эксперты в области когнитивных исследований говорят не просто о возможности формирования новых моделей игрового поведения посредством игр (о высшем уровне когнитивного взаимодействия), но и об экстраполяции этих моделей в реальную практику пользователей. Ярким примером здесь являются игры-симуляторы и VR-тренажеры [9].

Представим теперь компоненты видеоигры, определяя реализацию которых, гейм-дизайнер и команда разработчиков могут проектировать игровой опыт пользователей:

- цель (связана с обновлением опыта, в идеале – с приобретением новой модели поведения при взаимодействии с новой видеоигрой);

- воображаемый мир (сеттинг как возможность реализовать новую модель поведения в принципиально иной, по сравнению с реальным миром, среде, например более эстетичной или безопасной);
- воображаемая ситуация (игровой сюжет как возможность стать активным участником событий, новых для игрока);
- роль игрока (принятие им роли основного персонажа, его внешнего облика, характера, способностей и миссии);
- роли других участников (имитация социальных взаимодействий игрока с второстепенными персонажами);
- действия и взаимодействия (геймплей как ключевой системообразующий компонент видеоигры, определяющий эффективность когнитивного взаимодействия игрока с видеоигрой, наполнение и расширение его игрового опыта);
- средства игры (игровой интерфейс как способ взаимодействия с видеоигрой);
- правила (игровая механика как система условий и ограничений для реализации ранее сформированных моделей поведения, в том числе при взаимодействии с другими видеоиграми) [5].

При разработке указанных компонентов новой видеоигры необходимо понять, чем наполнен опыт ее потенциальных игроков на текущий момент. Для этого целесообразно провести UX-исследование (исследование опыта пользователей), направленное на выявление моделей поведения игроков, сформированных при взаимодействии с различными видеоиграми. Особенно полезными в этом контексте будут данные об игровом опыте пользователей прямых игр-аналогов.

Для организации такого исследования и структурирования его данных существуют различные фреймворки, которые подбираются в зависимости от жанра и первоначального варианта концепции новой видеоигры. К ним можно отнести, например, такой фреймворк, как MDA (англ., Mechanics – Dynamics – Aesthetics), фреймворки PENS (англ., The Player Experience of Need Satisfaction), SSM (англ., System – Story – Mental Model), DDE (англ., Design – Dynamics – Experience), Flow Theory, The 4 Keys 2 Fun и др. [10–13].

Кроме того в рамках методологии UX-исследования фреймворки можно сочетать с инструментами, которые часто применяются при исследовании опыта взаимодействия пользователей с неигровыми цифровыми продуктами, такими как карты персонажей и карты эмпатии, различные варианты JTBD-карт (англ., Jobs to be Done), CJM (англ., Customer Journey Map), карты ценностного предложения, Task Flow, User Flow и пр. [14, 15]. В рамках UX-исследования также стоит использовать различные онлайн-сервисы, которые автоматизируют визуализацию его данных, например, StoryBoardThat и пр. [16].

Представим более детальное описание фреймворка DDE, который был выбран нами для исследования игрового опыта потенциальных пользователей новой видеоигры. В соответствии с ним сбор, обработка, интерпретация и применение данных UX-исследования должны осуществляться по трем взаимосвязанным направлениям:

- 1) дизайн игры;
- 2) динамика игры;
- 3) опыт игрока, который включает в себя три уровня когнитивного взаимодействия с игрой:
 - ощущения (зависят от воздействия дизайна игры и ее динамики на органы чувств игрока, формируют восприятие отдельных компонентов игры);
 - церебеллум (эмоции и чувства игрока от игрового процесса, которые развиваются в устойчивые убеждения);
 - церебрум (осмысленная модель поведения игрока, которая вырабатывается как результат прохождения игры и получения игровых достижений) [8, 9].

Если дизайн игры и ее динамика в ходе UX-исследования устанавливаются как объективные данные, которые разработчик реализовал в конкретной игре, то компоненты «опыта игрока» в DDE описываются как субъективные данные, влияющие и на выбор новой видеоигры, и на степень вовлеченности в ее игровой процесс.

КОНЦЕПЦИЯ НОВОЙ ВИДЕОИГРЫ «КОЩЕЙСКИЕ ПРОДЕЛКИ»

Новая ролевая видеоигра «Кошечские проделки» разрабатывается небольшой командой инди-специалистов как стартап-проект (см. рис. 1). Он стал

победителем в федеральном конкурсе «Студенческий стартап», который организует Фонд содействия инновациям РФ [17].

Видеоигра ориентирована на русскоязычных пользователей, преимущественно подростков в возрасте 9–14 лет (основной профиль), а также на пользователей 17–23 лет (второстепенный профиль). Это игроки, которые любят фэнтезийные миры и захотят совместить для себя развлекательный контент с информацией о восточнославянской культуре, изложенной посредством новой интерпретации сказок, любимых с детства. Среди них можно выделить два психотипа по классификации Р. Бартла: «исследователи» (67%) и «карьеристы» (33%) [18].

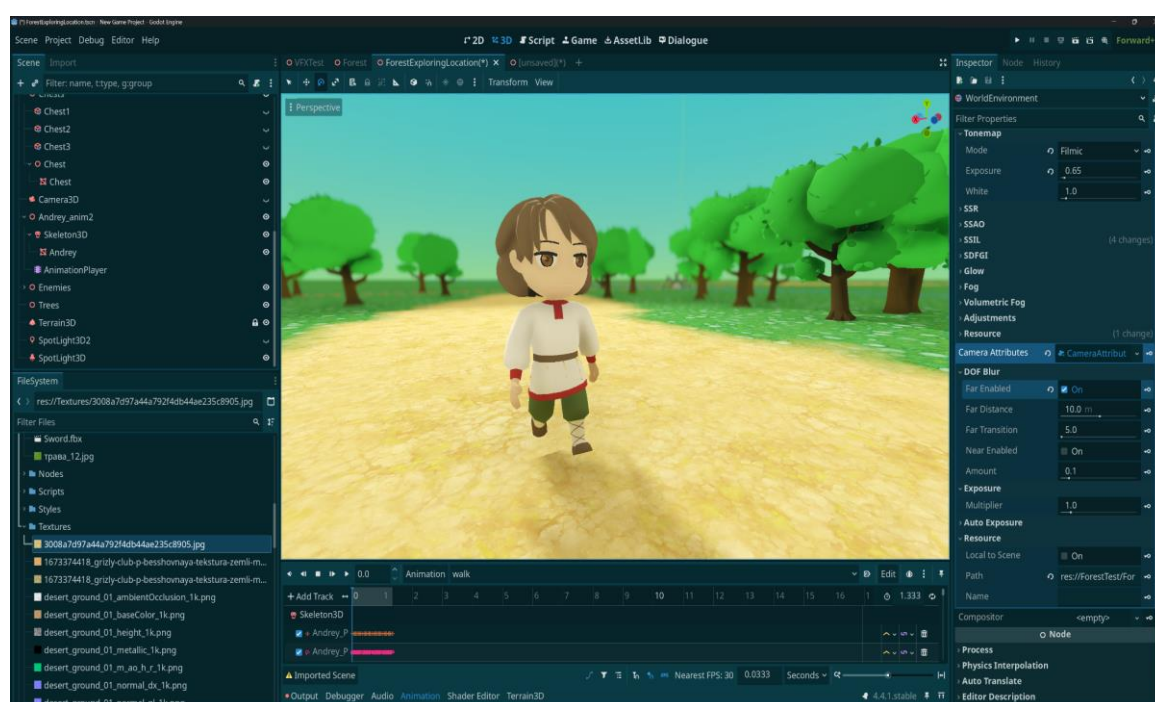


Рис. 1. Новая видеоигра «Кошечьи проделки»: главный персонаж

В основу сюжета видеоигры «Кошечьи проделки» заложены мотивы русских народных сказок. По сюжету, молодые и храбрые женихи, чьих невест на протяжении долгого времени похищал Кощей, собираются вместе, чтобы одолеть его. Им предстоит путешествовать по миру древней Руси, сражаться и заручаться поддержкой сказочных персонажей, таких как Баба Яга, Чудо-Юдо, Змей Горыныч, Верлиока и др., которые владеют информацией, важной для продвижения игроков, и могут помочь им со снаряжением. Постепенно локальная история о парнях из русских деревень перерастет для игроков в эпик о борьбе

добра со злом: им придется не просто выкрасть своих собственных невест, но и выполнить миссию от Царевича – победить в бою Кощея и освободить Русь от его «темных проделок».

Пространство игры реализуется по типу «хаб», где опорная локация предназначена для получения новых заданий и информации от второстепенных персонажей, создания отряда, покупки и выбора снаряжения его участникам. Из хаба игрокам будут доступны такие походные локации, как «Русский лес», «Черные горы», «Змеиная пещера», «Бездонное море» и, конечно, «Замок Кощея»: в них можно будет собирать целебные ягоды с кустов, находить секретные зоны со сказочными артефактами, вскрывать кощейские сундуки, а также решать головоломки при помощи волшебных заклинаний и сражаться с сильными и могущественными врагами (см. рис. 2).

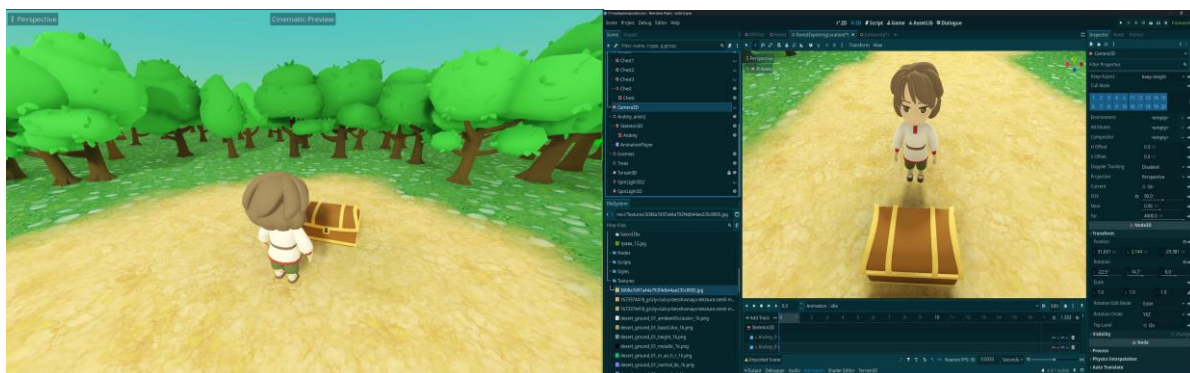


Рис. 2. Новая видеоигра «Кощейские проделки»: локация «Русский лес»

Разработка видеоигры осуществляется на движке Godot, функциональные возможности которого позволяют реализовать запланированное для нее 3D-пространство, геймплей, интерактивные диалоги персонажей и аудиовизуальные эффекты. Его преимущество состоит в наличии открытого исходного кода, что значительно уменьшает вероятность его запрета на территории РФ.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ ИГРОВОГО ОПЫТА ПОЛЬЗОВАТЕЛЕЙ

Исследование игрового опыта и формирование когнитивной модели взаимодействия пользователей с новой видеоигрой «Кошечьи проделки» произведены с использованием фреймворка DDE и карты пути пользователя (CJM). Авторская методология исследования базируется на парсинге и контент-анализе вторичных данных из открытых источников в интернете.

Опишем процедуру UX-исследования.

Этап 1. Выбор прямого аналога, схожего по различным компонентам игрового процесса с новой видеоигрой: если UX-исследование проводится на стадии формирования ее концепции, то аналогом может быть не только уже существующая видеоигра, но и нецифровой игровой продукт, например настольная карточная или спортивная командная игра. Кроме того, в качестве аналога может быть выбраны мультфильм, сериал, комикс или литературное произведение (последние варианты подходят тогда, когда в основе новой видеоигры есть запланированная сюжетная линия или конкретный нарратив).

Этап 2. Формирование перечня гипотез: в зависимости от того, на какой стадии разработки игры осуществляется UX-исследование, это могут быть гипотезы о проблемах (например, отсутствие данных об оптимальной степени сложности квестов для пользователей новой видеоигры или наличие данных о том, что треть пользователей бросает прохождение игры, когда доходит до определенного уровня, и др.), либо гипотезы о решениях (например, какие из вариантов перемещения главного персонажа, возможных к реализации в игре, помогут пользователям продвигаться по сюжету – ходьба, бег, бег с ускорением, прыжки из точки А в точку В или двойные прыжки и др.). Оба варианта гипотез могут быть сформулированы как утверждение или вопрос. Основной принцип их формирования базируется на когнитивном характере взаимодействия пользователя с игрой: чем больше гипотез будет сформулировано, тем больше полезных данных для разработки новой видеоигры будет получено в результате UX-исследования.

Этап 3. Систематизация гипотез UX-исследования по блокам, позволяющим оптимизировать сбор и последующую обработку данных в интернете. На этом этапе стоит воспользоваться одними из ранее описанных фреймворков

или карт, которые имеют готовую информационную структуру. Гипотезы в каждом блоке можно дополнить градацией эмоций или предпочтений (оценок) пользователей, например, дифференцировать их на позитивные, негативные и нейтральные. Такая систематизация позволит не выходить за пределы цели UX-исследования и организовано, не увлекаясь сбором и анализом всех упоминаний игры-аналога, парсить только те данные, которые полезны для понимания игрового опыта потенциальных пользователей новой видеоигры, значит, помогут при ее разработке.

Этап 4. Создание формы для фиксации и обработки данных, собранных в ходе UX-исследования. Для этого удобно использовать программы типа GoogleForms, YandexForms и пр.

Этап 5. Сбор данных в интернете и обработка результатов UX-исследования: чтобы получить объективные результаты UX-исследования, подтвердить или опровергнуть сформулированные гипотезы, целесообразно найти 2–3 источника информации на каждую из них – основной и контрольные. Источниками вторичных данных об игровом опыте пользователей могут быть форумы, специализированные площадки для продажи игр и стриминговые сервисы, социальные сети, где публикуются обзоры игр, а также отзывы и комментарии к ним, сайты консалтинговых компаний, размещающих результаты ранее проведенных специалистами опросов, и пр.

Для фиксации вторичных данных из интернета в созданную ранее онлайн-форму необходимо внести источники данных (названия и ссылки на доступ), информацию об актуальности данных (период их создания и появления в интернете), информацию о выборке или охвате аудитории (например, в выборку UX-исследования вошли 324 отзыва, которые игроки оставили об аналоге за последние 2 месяца), а также о методах и инструментах их обработки.

Обработку данных UX-исследования стоит производить по следующим критериям:

- подтверждают гипотезу;
- частично подтверждают гипотезу;
- опровергают гипотезу;
- частично опровергают гипотезу;

- новая информация, выходящая за пределы выдвинутых гипотез;
- информация по гипотезе в открытых источниках не найдена.

Этап 6. Визуализация результатов UX-исследования: создав легенду цветовых обозначений для перечисленных критериев обработки данных, необходимо внести в карту или фреймворк с гипотезами соответствующую цветовую дифференциацию. Это значительно упростит восприятие информации об игровом опыте потенциальных пользователей новой видеоигры и позволит корректно выполнить заключительный этап UX-исследования.

Этап 7. Интерпретация результатов UX-исследования и разработка рекомендаций по формированию когнитивной модели взаимодействия пользователей с новой видеоигрой.

Гипотезы, которые подтвердились вторичными данными, становятся основой для формирования когнитивной модели взаимодействия пользователей с новой компьютерной видеоигрой: она может быть описательной либо визуализированной с помощью схемы игрового цикла, Flow map, Mind map и других инструментов проектирования.

Гипотезы, которые частично подтвердились либо информация по которым не была найдена в открытых источниках в интернете, включаются в работу над новой видеоигрой и, как правило, проверяются на второй стадии UX-исследования с помощью различных методов сбора первичных данных, например опроса потенциальных игроков или тестирования прототипа.

Гипотезы, которые не подтвердились, чаще всего удаляются из итоговых фреймворков и карт, чтобы облегчить визуальное представление и последующее применение данных UX-исследования.

Новая информация, которая не сопоставима ни с одной из гипотез, но при этом была выявлена в ходе анализа отобранных вторичных данных, является наиболее ценной для уточнения концепции новой видеоигры, так как чаще всего содержит инсайты от потенциальных игроков относительно недостатков в имеющихся аналогах.

РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ ИГРОВОГО ОПЫТА ПОЛЬЗОВАТЕЛЕЙ

Представим промежуточные результаты UX-исследования, которое продолжается на текущий момент для увеличения выборки и повышения репрезентативности данных.

В качестве аналога новой видеоигры «Кошечьи проделки» была выбрана японская ролевая и приключенческая игра Persona 5 Royal с большим охватом аудитории, вышедшая в 2019 г. (см. рис. 3). Так как команда разработчиков планирует использовать ее геймплей практически в неизменном виде, интегрировав его с сеттингом русских сказок, персонажами из славянского фольклора и новым нарративом, важно понять, действительно ли такой подход понравится потенциальным игрокам и обеспечит коммерческий успех их стартап-проекту.



Рис. 3. Видеоигра Persona 5 Royal

Для получения информации об игровом опыте пользователей Persona 5 Royal было сформулировано 117 гипотез в форме утверждений, которые распределялись по двум профилям игроков – с психотипом «исследователь» и психотипом «карьерист», а затем поочередно заносились в несколько фреймворков DDE (см. рис. 4).

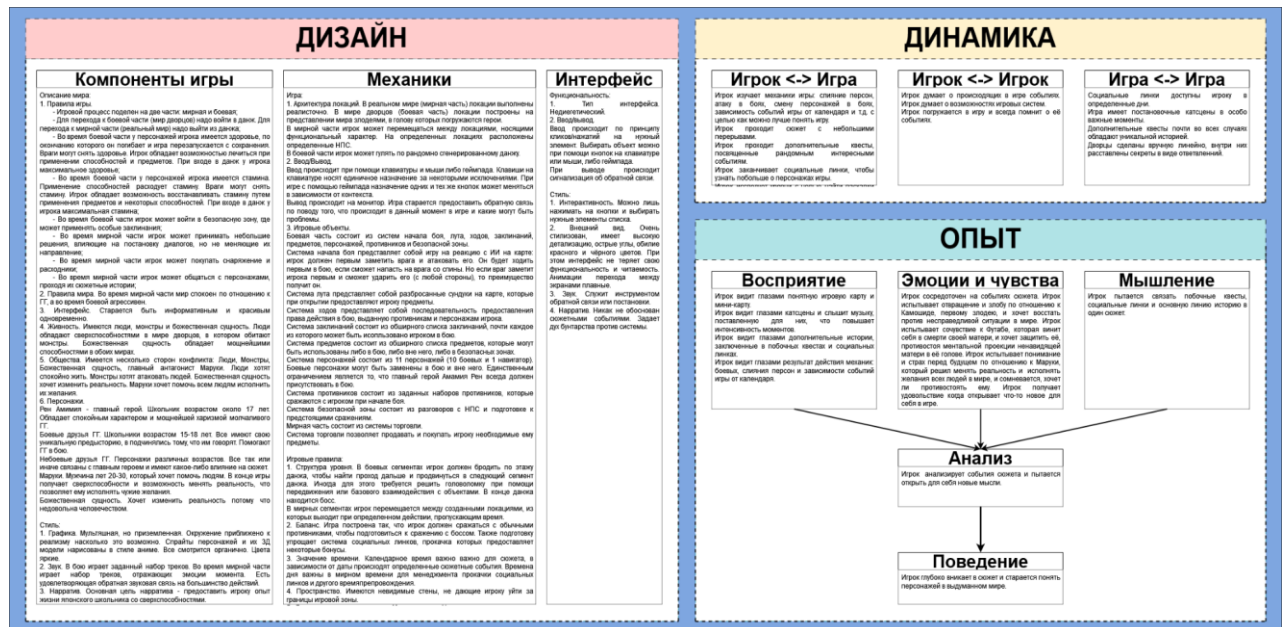


Рис. 4. Фреймворк DDE с гипотезами об игровом опыте пользователей Persona 5 Royal

В каждом фреймворке DDE в блоки «Дизайн» и «Динамика» заносились данные по компонентам игры Persona 5 Royal, которые фактически реализованы разработчиками и были зафиксированы авторами в процессе ее прохождения, а в блок «Опыт» – предположения относительно их восприятия, эмоциональной оценки, осмысления и поведения со стороны игроков.

Далее по каждой гипотезе был осуществлен парсинг информации в открытых источниках интернета. Для ее фиксации была создана Google-таблица. В нее заносились вторичные данные по следующим категориям UX-исследования: выборка, примеры отзывов, источник данных и ссылка на него, вывод относительно подтверждения или опровержения гипотезы, наличие или отсутствие новой информации об игровом опыте пользователей (см. рис. 5).

docs.google.com/spreadsheets/d/1CvEhpeeWkiBFQon64RXS1EXlqC2JKZ7bdpVLviro3qY/edit?gid=0#gid=0					
UX-исследование					
Файл Правка Вид Вставка Формат Данные Инструменты Расширения Справка					
Меню 100% 123 По ум...					
PERSONA 5 ROYAL					
Выборка	Гипотеза	Данные	Ссылка	Вывод	Новая информация
Исследователь: Задача 1					
Отзывы игроков: 38 Комментарии игроков: 19	Игрок выделяет пару часов свободного времени для изучения нового участка дворца. Вероятно, в выходной день.	Это может звучать не как совет, но не бойтесь проходить дворцы до конца за одно посещение, хоть и можно не спать, но лично я, когда проходил до самого конца, пока возможно. Чем раньше закончите путь до сокровища и украдете сокровище, больше времени будет на остальное.	https://steamcommunity.com/profiles/76561198140110213/recommended/1687920/	Подтверждена только часть про время.	Отсутствует
	Он полон сил, сидит за компьютером и полностью сосредоточен на задаче.	Palaces can take a decent amount of time to complete depending on how resource friendly you decide to be when it comes to items and SP management, but they can be doable within one day (unless the story forces you out to come back later for plot reasons).	https://steamcommunity.com/id/bo-nafidemonafide/recommended/1687920/		
Отзывы игроков: 17 Комментарии игроков: 31	Игрок в роли "Благородного вора сериан" видит, как реалит себя человек в реальной жизни (как очень некорректный человек если это коррупционер-тиран, или как человек, который не может справиться со своими травмами самостоятельно, если это более положительный злодей). Чтобы заставить человека измениться в лучшую сторону, игрок переносится в его дворец — по сути внутренний мир. В этом дворце игрок должен сделать "тень" человека до наступления определенной игровой даты. При этом, в связи с антагонизмом злодеев или травмами людей, "тени" будут препятствовать исследованию: искать игрока и пытаться победить его.	While chemistry between the characters never quite reaches the heights of Persona 4's Scooby Gang vibe, Persona 5's more cohesive main story does allow each character to shine in individual moments instead, exploring troubled pasts like the loss of a parent to add weight to present day scenarios. This series has never shied away from exploring serious subject matter, and here it shines light on abuse by teachers, drug sales, suicide, and a host of other issues that are handled well and are at times extraordinarily moving. Each character's motivation for joining you is relatable and unique, and even each palace's target has believable conviction. An unrepentant drug dealer, for example, reveals an origin that doesn't necessarily justify his crimes, but fleshes him out into more than just a bad guy caricature.	https://www.wign.com/articles/2017/03/28/persona-5-review-ew/		
		Persona's deep, thematic exploration of human society and the perils of the psyche continues to ask challenging and relevant questions about justice and suffering that left me rethinking my own convictions — and it's when a game isn't afraid to go to those places that it really becomes a work of higher art.	https://www.wign.com/articles/persona-5-royal-review/	Игрокам это нравится	Отсутствует
		Those dungeons are excellent, by the way, and by far the biggest single point of improvement	https://www		

Рис. 5. Google-таблица с данными об игровом опыте пользователей Persona 5 Royal

После этого был выполнен контент-анализ: с его помощью обработано 672 отзыва и комментария на разных языках, а также 42 русскоязычных обзора по игре Persona 5 Royal.

В табл. 1 представлены основные результаты проведенного UX-исследования, которые получены относительно базовых уровней когнитивного взаимодействия игроков-«исследователей» с Persona 5 Royal.

Таблица 1. Восприятие игроками Persona 5 Royя ее различных компонентов, их эмоции и чувства

Игрокам нравится	1. Интригующее начало игры, которое задает тон всему ее дальнейшему повествованию. 2. Понятный и близкий многим нарратив из реальной жизни о несправедливости мироустройства и необходимости для обычных людей противостоять злодеям-тиранам из высшего класса общества. 3. Хороший туториал длительностью в несколько часов, реализованный в первой локации игры. 4. Внешность и характеры персонажей игры, каждый из которых имеет свою собственную историю: например, игрокам-«исследователям» сильно нравится Маруки, один из финальных боссов, который исполняет желания всех людей на Земле просто потому, что хочет сделать их счастливыми. 5. Проработанный на контрастах сеттинг городского фэнтези, в котором потусторонний и реальный миры противопоставлены друг другу: в потустороннем измерении игроки исследуют подсознательные и явные помыслы главных злодеев игры, а в реальном измерении – живут обычной жизнью. 6. Динамичный игровой процесс, в котором гармонично сочетаются механики «мирной жизни» обычного японского школьника с боевой системой, предназначенной для победы над тенями. Оба типа механик одинаково важны для игроков: им нравится их изначально неочевидная связка. Так, например, в реальном мире игроки-«исследователи» сначала замечают, что Мишима ведет себя странно, а потом только в потустороннем мире понимают, в чем причины такого поведения, и решают вступить в бой с его тенью. 7. Захватывающие пошаговые бои, в которых все персонажи подконтрольные игроку, обладают детально раскрытым характером, предысторией, а главное – уникальными навыками, что позволяет игрокам-«исследователям» экспериментировать.
------------------	---

	8. Возможность выбрать и пройти за бонусы дополнительную сюжетную линию во время мирной жизни персонажа. 9. Неожиданный для игрока поворот под конец игры, завязанный на особенностях сеттинга: персонажи игрока умело обманывают всех (включая его самого) путем незаметной подмены реального мира его потусторонней копией.
Игрокам не нравится	1. Затянутость некоторых сюжетных сцен, например сцены в конце игрового ноября, когда игроку посредством двух диалогов в течение часа реального времени объясняют сложный план персонажей. 2. Вторичность отдельных сюжетных событий, например возвращения в команду Морганы или посещение дома художника Юскэ.
Игроки чувствуют	1. Принимая роль главного персонажа, игрок постепенно начинает чувствовать себя крутым, смелым и умным вором, который втайне исследует мир злодеев и способен решить все его головоломки: когда у него все получается – он всемогущ. 2. По мере раскрытия сюжета у игрока возникают контрастные чувства: перед глобальной несправедливостью мира он, наоборот, бессилен. 3. Напряжение для игрока с каждым новым событием нарастает: он испытывает все большую неприязнь, ненависть и гнев к злодеям, например к Камошиде, а к обычным людям – жалость. Так, например, игрок очень сочувствует Футабе, которая винит себя в смерти своей матери, и потому хочет защитить ее. 4. Яркими чувствами для игрока также являются страх и сомнение, например, относительно противостояния злодею Маруки, который исполняет желания всех людей в мире, чтобы сделать их счастливыми. 5. Каждый раз, когда игрок открывает для себя новое противоречивое чувство от лица главного персонажа, находя в

	нем собственное отражение, он получает огромное удовольствие, так как исследует таким способом свой собственный внутренний мир.
--	---

В табл. 2 представлены результаты проведенного UX-исследования, которые получены относительно высшего уровня когнитивного взаимодействия игроков-«исследователей» с Persona 5 Royal.

Таблица 2. Основные паттерны взаимодействия игроков с Persona 5 Royal

Цель игроков	Разобраться в причинах конфликта разных сторон
Мотивация игроков	Игрок в роли «благородного вора сердец» хочет заставить человека из реальной жизни измениться в лучшую сторону: для этого он должен одолеть его «тень».
Механики, которые у игроков ассоциируются с достижением цели	– передвижение по дворцу; – использование укрытий, чтобы избегать сражений; – заскриптованное взаимодействие с объектами; – решение головоломок путем сопоставления информации, полученной из дворца и реального мира; – использование конкретных предметов, например «Дымовой завесы», чтобы стать невидимым для «теней», или «Успокаивающего аромата», чтобы уменьшить бдительность врагов; – использование внутриигрового календаря для контроля за временем и определения доступности персонажей в реальном мире; – просмотр катсцен (от англ. «cut scene») о конфликте между героями и злодеями, их связи между собой, а также связи между реальным миром и главным злодеем.

Компоненты сеттинга, которые больше всего запомнились игрокам	1. Дворец как внутренний мир человека (обычно злодея): вначале свобода передвижения по нему ограничена, но постепенно расширяется по ходу сюжета. Исследуя новые части локации в определенном разработчиками порядке, игрок постепенно осознает, как владелец дворца видит мир и людей вокруг себя. 2. Сундуки: их игрок может найти при исследовании и получить полезные для себя снаряжение или предметы. 3. Секретные зоны во дворцах: в них игроку доступна тайна чувств, которые руководят владельцами дворцов.
Игрокам помогают	– обучение в первом дворце, которое сопровождается множеством всплывающих окон с подсказками; – мини-карта для навигации; – подсказки о текущей миссии, выполнение которой необходимо для открытия следующих зон во дворцах; – подсказки о возможности взаимодействия с объектами, а также укрытия за ними; – реплики персонажей, которые подсказывают игроку, какие действия необходимы для решения головоломки; – звуки, сигнализирующие игроку, что рядом находится сундук или секретная зона; – предметы, которые усыпляют бдительность врагов и тем самым позволяют продолжить исследование.
Игрокам мешают	– некоторые головоломки, которые можно решить только методом «проб и ошибок»; – автоматическое выкидывание игрока из дворца, когда уровень его тревоги повышается в случае обнаружения врагами.

Внешний контекст для игроков	Среднестатистический игрок выделяет в среднем пару часов в выходной день для изучения нового участка дворца. Это требует от него сосредоточенности.
Достижения игроков	1. Дойти до дворца главного злодея. 2. Победить мини-боссов «Стражей». 3. Найти и получить все Семена грехов во всех 9 дворцах злодеев.

Для наглядности и удобства применения результатов исследования игрового опыта пользователей-«исследователей» выбранного нами аналога Persona 5 Royal, которые детально описаны в табл. 1 и 2, визуализируем их с помощью типового шаблона CJM (см. рис. 6).

 Задача Найти x9							
Механики	Передвижение	Укрытие	Взаимодействие с объектами	Головоломки	Использование предметов	Внутриигровой календарь	Просмотр катсцен
Эмоции							
Оценка							

Рис. 6. CJM игрока-«исследователя» по Persona 5 Royal

КОГНИТИВНАЯ МОДЕЛЬ ВЗАИМОДЕЙСТВИЯ ПОЛЬЗОВАТЕЛЕЙ С НОВОЙ ВИДЕОИГРОЙ «КОЩЕЙСКИЕ ПРОДЕЛКИ»

Сформируем когнитивную модель взаимодействия пользователей с новой видеоигрой «Кошнейские проделки» путем экстраполяции части данных, которые были получены в результате исследования игрового опыта пользователей Persona 5 Royal. В табл. 3 описание ее основных аспектов.

Таблица 3. Основные аспекты когнитивной модели взаимодействия игроков с новой видеоигрой «Кощейские проделки»

У игроков будет цель	Разобраться в конфликте между сторонами добра и зла – миром простых и открытых людей из русских деревень и коварным Кощеем с его свитой.
Игроков будет мотивировать	Побыть в роли простого и смелого парня, обладающего настоящей «русской смекалкой», который ради спасения своей любимой невесты готов обойти всю сказочную Русь и найти место, где ее спрятал Кощей. По мере развития сюжета его роль станет более глобальной: как настоящий герой русских сказок он возьмет на себя миссию освобождения Руси от гнета Кощея.
Для достижения цели игроки будут делать	– возвращаться в локацию-хаб – «Избушку Бабы Яги»; – выбирать помощников для похода и назначать им снаряжение: например, игрок сможет выбрать для похода в «Русский лес» Ваню-дурачка, чтобы сразиться с Лешим, и назначить ему деревянный меч, кольчугу и шапку-невидимку; – передвигаться по походным локациям, таким как «Русский лес», «Черные горы», «Болото Кикиморы», «Змеиная пещера», «Бездонное море», «Целебное озеро», «Новгород» и «Замок Кощея»; – взаимодействовать с объектами (с единственным вариантом действия): например, когда игрок найдет корзину с волшебными яйцами, в одном из которых находится смерть Кощея, он сможет увидеть их ауру, нажимая на кнопку геймпада; – решать головоломки посредством зашифрованного взаимодействия с объектами: например, в вышеупомянутом примере игрок сможет по цвету ауры каждого яйца понять, в каком из них находится игла – Смерть Кощея;

	– использовать волшебные возможности сказочных предметов: например, таких как «Волшебная метла», которая мгновенно возвратит игрока в локацию-хаб, когда у него закончатся ресурсы на поход, или «Молодильное яблоко», которое залечит рану персонажа-помощника после сражения; – сражаться, подбирая правильные магические приемы под слабости врагов: например, против Лешего в локации «Русский лес» игрок сможет применить сильное заклинание «Чтоб тебе пусто было»; – просматривать катсцены о между сторонами добра и зла, их связи между собой: например, о конфликте между Кощеем и Бабой Ягой – невестой, сумевшей сбежать от Кощея прямо со свадьбы, которую он затем превратил в старуху.
Игрокам помогут продвигаться по игре	– локация «Русский лес», предназначенная для обучения и адаптации к механикам игры: в ней игровой процесс будет сопровождается множеством всплывающих окон с подсказками по управлению главным персонажем; – мини-карта навигации по миру Руси; – подсказки о текущих заданиях в книге русских сказок (интерфейсное решение), выполнение которых необходимо для продвижения по сюжету; – визуальные подсказки о возможности взаимодействия с объектами; – аудиальные подсказки о возможности найти рядом сундук Кощея; – реплики персонажей, которые по запросу игрока подскажут ему решение головоломки; – предметы, созданные из найденных игроком материалов: например, сосуд с зельем, которым можно усыпить любого врага из свиты Кощея;

Игроки запомнят	– походные локации как места скопления зла; – хаб как место силы главного героя и его помощников; – команду главного героя – персонажей-помощников, чьи способности дополняют друг друга; – злых персонажей из русских сказок как отражение сути обычных людей с большим количеством пороков, но при этом не утративших до конца свою человечность; – Бабу-Ягу и Кощея как олицетворение тонкой связи между миром добра и миром зла, возможности выбрать и перейти на любую из сторон конфликта; – секретные зоны походных локаций, в которых игрок откроет для себя обстоятельства, из-за которых их обитатели перешли на сторону зла; – волшебные предметы как уникальный способ победить злодеев из русских сказок.
Игроки почувствуют и сформируют убеждение	1. Базовой для игрока-«исследователя» эмоцией в игре будет любопытство, постоянное поддержание которого будет обеспечиваться за счет новой интерпретации сюжета русских сказок, а также придуманной предыстории злодеев из свиты Кощея. 2. Также на протяжении игрового процесса у него будут развиваться чувство автономности (свободы действий в сочетании с ответственностью), актуальное для психотипа «исследователь», а также чувство значимости. Так, например, игрок станет лидером команды добрых и смелых парней из русских деревень: по ходу игрового процесса он вместе с ними разберется в волшебных сущностях и способностях Кощейской свиты, поняв, что причина любого кроется в мире обычных людей. 3. Убеждением игрока, которое закрепится в его опыте на выходе из игры, станет следующий нарратив: добро и зло

	живут в сердце абсолютно каждого человека, поэтому не дать жизненным обстоятельствам возобладать над собой и остаться на стороне добра – это главная победа.
Игроки получают достижения	1. Откроют 8 походных локаций, исследуют их территорию и обнаружат секретные зоны. 2. Найдут 8 сундуков Кощея и соберут из них снаряжение. 3. Победят 8 мини-боссов – злыдней из свиты Кощея, таких как Леший, Кикимора, Огненный змей, Чудо-Юдо, Змей Горыныч, Полуденица, Верлиока и Чёрт Кузнец. 4. Дойдут до замка Кощея, победят его и откроют темницу с похищенными невестами.

Визуализируем часть когнитивной модели, связанной с логикой действий игроков-«исследователей» новой разрабатываемой видеоигры «Кощейские проделки» (см. рис. 7).



Рис. 7. Логика действий игроков-«исследователей» новой видеоигры «Кощейские проделки»

ЗАКЛЮЧЕНИЕ

Методология, разработанная авторами, позволила детально исследовать игровой опыт пользователей *Persona 5 Royal*. Его результаты были применены для формирования когнитивной модели взаимодействия пользователей с новой видеоигрой «Кошечьи проделки».

Показано, как можно интегрировать логику поведения игроков популярного аналога с новым сеттингом и сюжетным нарративом, скорректировав таким образом концепцию новой видеоигры, находящейся на ранней стадии разработки.

Результаты UX-исследования подтвердили важность учета таких уровней когнитивного взаимодействия пользователей с видеоигрой, как восприятие, эмоции, чувства, убеждения, модели поведения и достижения, для обеспечения их глубокой вовлеченности в игровой процесс. Они также показали, что разработанная авторами методология исследования игрового опыта универсальна: она может быть использована и в других игровых проектах, в том числе для изучения трансформации игрового опыта в зависимости от жанровых и культурных предпочтений игроков.

Благодарности

Авторы выражают благодарность Федеральному государственному бюджетному учреждению «Фонд содействия развитию малых форм предприятий в научно-технической сфере» за поддержку стартап-проекта по разработке новой видеоигры «Кошечьи проделки».

ЛУДОГРАФИЯ

Atlus Co., Ltd. (2024) [2019.] *Persona 5 Royal* [видеоигра][загрузка, Blu-ray диск][PlayStation Vita, Nintendo Switch, PlayStation 4, PlayStation 5, Xbox One, Windows, Xbox Series X|S]. Sega Corporation.

СПИСОК ЛИТЕРАТУРЫ

1. Martinez L., Gimenes M., Lambert E. Video games and board games: Effects of playing practice on cognition. URL: <https://doi.org/10.1371/journal.pone.0283654>

2. *Pallavicini F., Ferrari A., Mantovani F.* Video Games for Well-Being: A Systematic Review on the Application of Computer Games for Cognitive and Emotional Training in the Adult Population.

URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6234876/>

3. *Sattar S., Iqbal Y., Yousaf R.* Impact of Playing Video Games on Cognitive Functioning and Learning Styles of Adolescents.

URL: https://www.researchgate.net/publication/357385738_IMPACT_OF_PLAYING_VIDEO_GAMES_ON_COGNITIVE_FUNCTIONING_AND_LEARNING_STYLES

4. *Macklin C., Sharp J.* Games, Design and Play: A Detailed Approach to Iterative Game Design. Boston: Addison-Wesley. 2016. 266 p.

5. *Zubek R.* Elements of Game Design. London: The MIT Press, 2020. 252 p.

6. *Dobrowolski P., Hanusz K., Sobczyk B., Skorko M., Wiatrow A.* Cognitive enhancement in video game players: The role of video game genre.

URL: <https://www.sciencedirect.com/science/article/abs/pii/S074756321400661X>

7. *Corso D.* A Review of Video Game Effects and Uses.

URL: https://sc.edu/about/offices_and_divisions/research/news_and_pubs/caravel/archive/2014/2014-caravel-video-games.php

8. *Caroux L., Isbister K., Bigot L., Vibert N.* Player video game interaction: A systematic review of current concepts.

URL: <https://www.sciencedirect.com/science/article/abs/pii/S0747563215000941>

9. *Shoshani A.* From virtual to prosocial reality: The effects of prosocial virtual reality games on preschool Children's prosocial tendencies in real life environments.

URL: <https://www.sciencedirect.com/science/article/abs/pii/S0747563222003661>

10. *Duarte L.* Revisiting the MDA framework. URL: <https://www.gamedeveloper.com/design/revisiting-the-mda-framework>

11. *Dillon R.* The 6-11 framework: A new methodology for game analysis and design.

URL: https://www.academia.edu/1571687/THE_6_11_FRAMEWORK_A_NEW_METHODODOLOGY_FOR_GAME_ANALYSIS_AND_DESIGN

12. *Fabricatore C., Nussbaum M., Rosas R.* Playability in action video games: A qualitative design model. URL: https://doi.org/10.1207/S15327051HCI1704_1

13. Walk W., Görlich D., Barrett M. Design, Dynamics, Experience (DDE): An Advancement of the MDA Framework for Game Design.

URL: https://www.researchgate.net/publication/315854140_Design_Dynamics_Experience_DDE_An_Advancement_of_the_MDA_Framework_for_Game_Design

14. Yee N. Motivations for Play in Online Games.

URL: <https://doi.org/10.1089/cpb.2006.9.772>

15. Rosenbaum M.-S., Otalora M.-L., Ramírez G.-C. How to Create a Customer Journey Map.

URL: https://www.researchgate.net/publication/309182735_How_to_create_a_realistic_customer_journey_map

16. Free Storyboarding Software – Online Storyboard Creator.

URL: <https://www.storyboardthat.com/>

17. Конкурс «Студенческий стартап».

URL: <https://fasie.ru/programs/programma-studstartup/>

18. Федоров А.В., Соколова И.И. Эмоциональное вовлечение в видеоиграх: роль нарратива и механик. URL: <https://doi.org/10.12345/exppsy.2020.567>

THE RESEARCH OF GAMING EXPERIENCE AND THE FORMATION OF A COGNITIVE MODEL OF USER INTERACTION WITH A NEW VIDEO GAME

M. V. Shmatko¹ [0000-0002-7255-8885], A. A. Kutuzov² [0000-0004-6652-2968],

L. R. Ponomarev³ [0009-0009-9600-7824]

^{1, 2, 3}*Omsk State Technical University, Omsk, 644050, Russia*

¹marin298@gmail.com, ²kutuzovschool@mail.ru, ³oil294semrom@gmail.com

Abstract

The article is devoted to the study of the gaming experience of users and the formation of a cognitive model of their interaction with the new video game "Koshchejskie prodelki", which is at an early stage of development. The UX research methodology proposed by the authors, as well as the results obtained during it, demonstrate how qualitative data on the perception, emotions, feelings, behavior

patterns and achievements of players help optimize the process of creating video games, making them not only entertaining, but also psychologically effective. The results of the work have practical value for indie developers who, with limited resources, strive to create high-quality gaming products.

Keywords: *video game, game design, game experience, user experience, cognitive interaction, cognitive model, UX research.*

LUDOGRAPHY

Atlus Co., Ltd. (2024) [2019.] *Persona 5 Royal* [видеоигра][загрузка, Blu-ray диск][PlayStation Vita, Nintendo Switch, PlayStation 4, PlayStation 5, Xbox One, Windows, Xbox Series X|S]. Sega Corporation.

REFERENCES

1. Martinez L., Gimenes M., Lambert E. Video games and board games: Effects of playing practice on cognition. URL: <https://doi.org/10.1371/journal.pone.0283654>
2. Pallavicini F., Ferrari A., Mantovani F. Video Games for Well-Being: A Systematic Review on the Application of Computer Games for Cognitive and Emotional Training in the Adult Population.
URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC6234876/>
3. Sattar S., Iqbal Y., Yousaf R. Impact of Playing Video Games on Cognitive Functioning and Learning Styles of Adolescents.
URL: https://www.researchgate.net/publication/357385738_IMPACT_OF_PLAYING_VIDEO_GAMES_ON_COGNITIVE_FUNCTIONING_AND_LEARNING_STYLES
4. Macklin C., Sharp J. Games, Design and Play: A Detailed Approach to Iterative Game Design. Boston: Addison-Wesley. 2016. 266 p.
5. Zubek R. Elements of Game Design. London: The MIT Press, 2020. 252 p.
6. Dobrowolski P., Hanusz K., Sobczyk B., Skorko M., Wiatrow A. Cognitive enhancement in video game players: The role of video game genre.
URL: <https://www.sciencedirect.com/science/article/abs/pii/S074756321400661X>
7. Corso D. A Review of Video Game Effects and Uses.
URL: https://sc.edu/about/offices_and_divisions/research/news_and_pubs/caravel/archive/2014/2014-caravel-video-games.php

8. *Caroux L., Isbister K., Bigot L., Vibert N.* Player video game interaction: A systematic review of current concepts.
URL: <https://www.sciencedirect.com/science/article/abs/pii/S0747563215000941>
9. *Shoshani A.* From virtual to prosocial reality: The effects of prosocial virtual reality games on preschool Children's prosocial tendencies in real life environments.
URL: <https://www.sciencedirect.com/science/article/abs/pii/S0747563222003661>
10. *Duarte L.* Revisiting the MDA framework. URL: <https://www.gamedeveloper.com/design/revisiting-the-mda-framework>
11. *Dillon R.* The 6-11 framework: A new methodology for game analysis and design.
URL: https://www.academia.edu/1571687/THE_6_11_FRAMEWORK_A_NEW_METHODODOLOGY_FOR_GAME_ANALYSIS_AND_DESIGN
12. *Fabricatore C., Nussbaum M., Rosas R.* Playability in action video games: A qualitative design model. URL: https://doi.org/10.1207/S15327051HCI1704_1
13. *Walk W., Görlich D., Barrett M.* Design, Dynamics, Experience (DDE): An Advancement of the MDA Framework for Game Design.
URL: https://www.researchgate.net/publication/315854140_Design_Dynamics_Experience_DDE_An_Advancement_of_the_MDA_Framework_for_Game_Design
14. *Yee N.* Motivations for Play in Online Games.
URL: <https://doi.org/10.1089/cpb.2006.9.772>
15. *Rosenbaum M.-S., Otalora M.-L., Ramírez G.-C.* How to Create a Customer Journey Map.
URL: https://www.researchgate.net/publication/309182735_How_to_create_a_realistic_customer_journey_map
16. Free Storyboarding Software – Online Storyboard Creator.
URL: <https://www.storyboardthat.com/>
17. Конкурс «Студенческий стартап» [Student startup competition].
URL: <https://fasie.ru/programs/programma-studstartup/>
18. *Fedorov A.V., Sokolova I.I.* Emocionalnoe вовлечение в видеоигrah: rol narrativa i mekhanik. URL: <https://doi.org/10.12345/exppsy.2020.567>

СВЕДЕНИЯ ОБ АВТОРАХ



ШМАТКО Марианна Владимировна – к. филос. н., доцент кафедры «Математические методы и информационные технологии в экономике» ОмГТУ. Сфера научных интересов – гейм-дизайн, UX-аналитика и юзабилити информационных систем. Сфера практической деятельности — цифровой маркетинг.

Marianna Vladimirovna SHMATKO – Candidate of Philosophical Sciences, Associate Professor of the Department «Mathematical Methods and Information Technologies in Economics» of Omsk State Technical University. Research interests: game design, UX analytics and usability of information systems. Practical activities: digital marketing.

email: marin298@gmail.com

ORCID: 0000-0002-7255-8885



КУТУЗОВ Андрей Андреевич – студент ОмГТУ, направление 09.03.02 «Информационные системы и технологии» (профиль – «Цифровые медиатехнологии, виртуальная и дополненная реальность»). Генеральный директор и гейм-дизайнер ООО «ИГРЫ ДЕЛАЮТ». Сфера научных интересов — разработка видеоигр, гейм-дизайн, игровая сценаристика.

Andrey Andreevich KUTUZOV – student of the Omsk State Technical University, majoring at 09.03.02 «Information Systems and Technologies» (specialization: «Digital Media Technologies, Virtual and Augmented Reality»). CEO and Game Designer of the «IGRY DELAYUT» company. Research interests: video game development, game design, and game narrative design.

email: kutuzovschool@mail.ru

ORCID: 0000-0004-6652-2968



ПОНОМАРЕВ Лев Романович – студент ОмГТУ, направление 09.03.02 «Информационные системы и технологии» (профиль – «Цифровые медиатехнологии, виртуальная и дополненная реальность»). Программист ООО «ИГРЫ ДЕЛАЮТ». Сфера научных интересов – игровой дизайн, игровые механики, игровая сценаристика, разработка видеоигр.

Lev Romanovich PONOMAREV – student of Omsk State Technical University, majoring at 09.03.02 «Information systems and technologies» (specialization: «Digital Media Technologies, Virtual and Augmented Reality»). Programmer of the «IGRY DELAYUT» company. Research interests: game design, game mechanics, game storytelling, video game development.

email: oil294semrom@gmail.com

ORCID: 0009-0009-9600-7824

Материал поступил в редакцию 12 мая 2025 года

УДК 004.946.2+004.932.72+004.272

СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК ДЛЯ ДЕТЕКТИРОВАНИЯ ПОЗЫ ЧЕЛОВЕКА В УСЛОВИЯХ РАБОТЫ НА МОБИЛЬНЫХ УСТРОЙСТВАХ

Е. И. Ярко^[0009-0009-0502-1066]

Казанский (Приволжский) федеральный университет, г. Казань, 420008,
Россия

yarkoeg@gmail.com

Аннотация

Оценка положения тела человека (Human Pose Estimation, HPE) стала одной из наиболее актуальных тем в исследованиях в области компьютерного зрения. Эта технология может применяться в различных сферах, таких как видеонаблюдение, медицинская помощь и анализ спортивных движений.

В связи с растущим спросом на HPE за последние 20 лет было разработано большое количество библиотек для этой технологии. С 2017 года опубликовано множество алгоритмов HPE, основанных на скелетной модели, которые были упакованы в библиотеки для удобства использования исследователями. Эти библиотеки важны для исследователей, которые хотят интегрировать их в реальные приложения для видеонаблюдения, медицинской помощи и анализа спортивных движений.

В работе рассмотрены преимущества и недостатки четырёх популярных передовых библиотек HPE для распознавания поз человека, которые могут работать на мобильных устройствах: Lightweight OpenPose, PoseNet, MoveNet и Blaze Pose.

Ключевые слова: поза человека, Human Pose Estimation, детектирование позы, компьютерное зрение, мобильные устройства, дополненная реальность, Lightweight OpenPose, PoseNet, MoveNet, BlazePose, скелетная модель.

ВВЕДЕНИЕ

Распознавание поз человека (Human Pose Estimation, HPE) – это процесс определения положения всех частей тела человека на входных изображениях или видео. В настоящее время технология HPE стала востребованной в сфере компьютерного зрения и дополненной реальности (Augmented Reality, AR). Она широко применяется в системах видеонаблюдения [1–6], медицинской помощи [7–15], анализе спортивных движений [16–28] и во многих других сферах. В частности, технологии определения позы человека успешно применяются в реабилитационных играх для постинсультных пациентов [47]. Такие игры позволяют сделать процесс реабилитации более эффективным и увлекательным для пациентов.

Для классификации поз и оценки их корректности в областях использования, перечисленных выше, используются ключевые точки на теле человека (рис. 1). С помощью интеллектуальной системы компьютерного зрения можно определить ключевые точки на скелете, что может быть применимо для решения различного спектра задач.

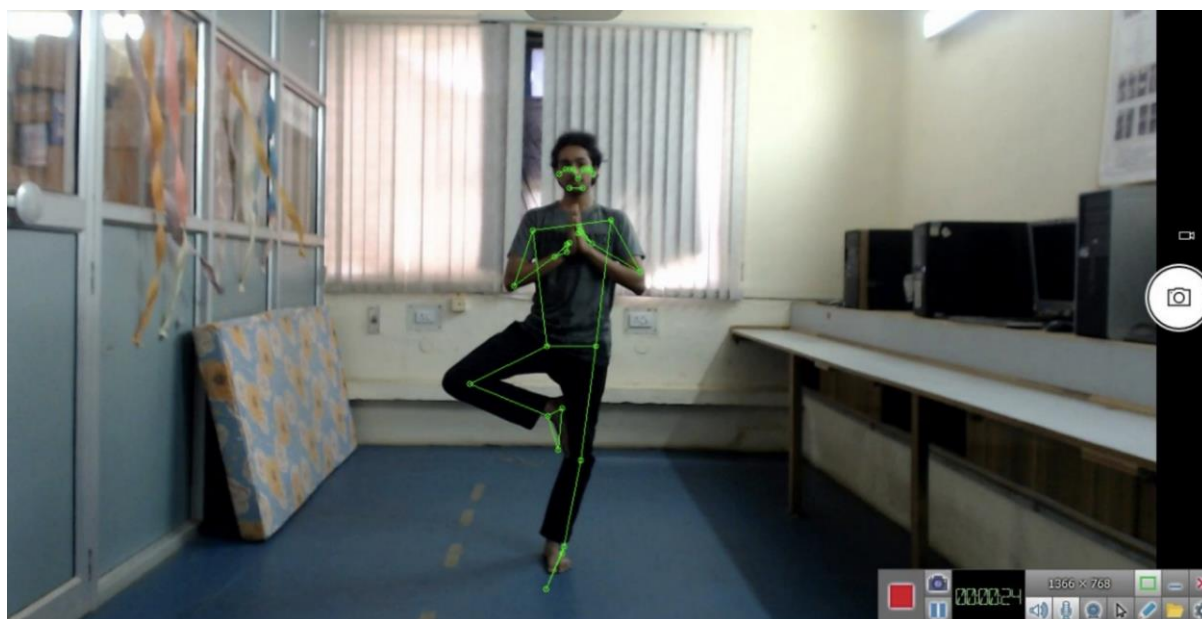


Рис. 1. Построение ключевых точек с помощью BlazePose

В сфере медицинской помощи детектирование ключевых точек на теле может быть использовано для оценки правильности выполнения упражнений

физиотерапии, обнаружения падений и реабилитации на дому. Кроме того, анализ ключевых точек на теле спортсмена позволяет оценить его производительность и корректность движений по сравнению с эталонными позами (ground truth).

HPE можно разделить на два типа: двумерную (2D) и трёхмерную (3D). Ключевое отличие между 2D- и 3D-оценками позы заключается в наличии дополнительной z-координаты, которая отвечает за глубину. Данные о глубине могут включать несколько камер (стерео- или мультивидовые системы), данные от глубинных датчиков (например, RGB-D-камеры) или использовать методы, которые восстанавливают глубину на основе одного RGB-изображения (моно-сетап), опираясь на статистическую регрессию.

Кроме того, HPE можно классифицировать по количеству определяемых в кадре людей: одного человека или нескольких. В обоих случаях, HPE можно дополнительно разделить на методы сверху вниз (top-down) и снизу вверх (bottom-up) в зависимости от способов детектирования ключевых точек скелета [29].

Основная задача настоящей работы – изучить практики применения библиотек детектирования позы человека с целью выбора оптимального решения в условиях работы программного обеспечения на мобильных устройствах, не обладающих глубинными датчиками. Мы представим сравнительное исследование двумерной оценки поз одного человека.

С ростом спроса на HPE было разработано большое количество алгоритмов, основанных на скелетной модели, которые объединены в библиотеки для удобства использования. Производительность этих библиотек HPE важна для обеспечения надёжности различных практических приложений, в которые они интегрируются. Например, если библиотека HPE используется в системе реабилитации на дому, она должна точно определять позы пациентов, выполняющих реабилитационные упражнения в различных домашних условиях, чтобы гарантировать надёжность приложения. Эта задача становится ещё более сложной, когда выявление ключевых точек скелета затруднено такими распространён-

ными проблемами, как неправильное положение камеры и самозакрытие (явление, при котором одна часть тела закрывается другой, что затрудняет детекцию, например, когда человек закрывает рукой лицо) [11, 20, 22, 23].

В последние годы четыре библиотеки HPE нашли применение в различных областях: PoseNet [30], MoveNet [31], Lightweight OpenPose [32] и Blaze Pose [33]. Во всех вышеприведенных библиотеках HPE возникают распространенные проблемы в оценке поз, такие как неправильное положение камеры и эффект самозакрывания.

Хотя каждая библиотека HPE использует свои подходы для преодоления отмеченных проблем, плюсы и минусы этих четырёх библиотек HPE пока не были исследованы. Поэтому необходимо провести сравнительный анализ производительности этих библиотек, чтобы оценить их надёжность в детектировании различных поз человека. Итак, цель этой работы – сравнить четыре передовые библиотеки HPE для распознавания поз человека и проанализировать их преимущества и недостатки.

2. СРАВНИТЕЛЬНЫЙ АНАЛИЗ БИБЛИОТЕК HPE

2.1. Существующий сравнительный анализ

В работах [33, 34], посвящённых сравнительному анализу, сопоставлена эффективность нескольких библиотек HPE, работающих с наборами изображений.

В одном из исследований [34] для сравнения OpenPose и BlazePose [33] был использован набор данных AR. В качестве критерия оценки была применена метрика PCK@0.2. Результаты показали, что OpenPose продемонстрировала несколько более высокую производительность по сравнению с BlazePose, набрав 87,8 и 84,1 балла соответственно.

В другом исследовании [31] были использованы OpenPose, PoseNet, MoveNet Lightning и MoveNet Thunder. Выбраны два набора данных изображений: COCO [35] и MPII [36]. Эффективность библиотек HPE была оценена с помощью собственной метрики. Наилучшие результаты показала PoseNet, в то время как MoveNet Lightning продемонстрировала худшие результаты.

Еще одно исследование [38] было посвящено сравнению четырех HPE библиотек: OpenPose, PoseNet, MoveNet и MediaPipe Pose. Авторы сравнили библиотеки на наборах данных из видеороликов и статичных изображений, а в качестве оценки производительности использовали показатель процентного охвата суставов (Percentage of Detected Joints, PDJ). MoveNet показала наилучшую производительность в выявлении различных поз человека на статических изображениях и видео. В [38] описаны проведенные эксперименты с использованием наборов данных изображений и видео.

2.2. Детальное сравнение библиотек HPE

В этом разделе рассмотрены библиотеки HPE, которые могут функционировать на мобильных устройствах, и их характеристики сведены в табл. 1.

Таблица 1. Характеристики библиотек оценки позы человека

Название библиотеки	Год выпуска	Максимальное число ключевых точек	Расположение ключевых точек в частях тела	Метод определения	Нейросетевая архитектура
PoseNet 2.0	2022	17	Глаза, уши, плечи, локти, запястья, бедра, колени, лодыжки	Bottom-up	ResNet
MoveNet	2021	17	Глаза, уши, плечи, локти, запястья, бедра, колени, лодыжки	Top-down	MobileNetV2
Lightweight OpenPose	2018	18	Включает шею, нос и дополнительные суставы рук	Bottom-up	MobileNet
BlazePose	2020	33	Полное тело с пальцами рук и ног	Top-down	BlazePose CNN

В библиотеках HPE было выявлено 17 ключевых точек, которые встречаются наиболее часто. Среди них пять точек расположены на голове: это уши,

глаза и нос. Шесть ключевых точек находятся на плечах, локтях и запястьях и относятся к верхней части тела, а шесть точек на бёдрах, коленях и лодыжках – к нижней части тела. Однако в Lightweight OpenPose добавлена одна дополнительная ключевая точка на шее, а BlazePose – более детализированные точки на лице и руках (например, мочки ушей, кончики пальцев).

В библиотеках НРЕ можно выделить два подхода к обнаружению ключевых точек: «сверху вниз» (Top-down) и «снизу вверх» (Bottom-up).

В подходе «сверху вниз» сначала определяют количество людей на изображении, после чего каждому человеку присваивают отдельную рамку. Далее в каждой рамке происходит обнаружение ключевых точек. В отличие от подхода «сверху вниз», в подходе «снизу вверх» сначала устанавливаются ключевые точки, а затем они группируются по человеческим фигурам.

Среди рассматриваемых библиотек подход «сверху вниз» используют MoveNet и BlazePose, тогда как PoseNet 2.0 и Lightweight OpenPose – подход «снизу вверх».

PoseNet 2.0 использует свёрточные нейронные сети MobileNetV1 и MobileNetV2 для эффективной оценки позы на мобильных устройствах и в браузерах. MobileNet обладает легкой архитектурой, что позволяет сети работать быстро и с низкими требованиями к вычислительным ресурсам, сохраняя точность распознавания позы на базовом уровне.

Библиотека MoveNet представлена в двух вариантах: MoveNet Lightning и MoveNet Thunder. MoveNet Lightning работает на основе MobileNetV3, обеспечивая высокую производительность на устройствах с ограниченными ресурсами, MoveNet Thunder использует более крупную архитектуру, разработанную для высокой точности на более мощных устройствах. Эти сети оптимизированы для точной и быстрой работы в режиме реального времени.

Сеть Lightweight OpenPose была разработана с использованием облегченных версий сетей MobileNet и VGG. В этой версии OpenPose применены упрощенные и сжимаемые сети для обеспечения более эффективной работы на устройствах с низкими вычислительными возможностями, с сохранением при этом основной структуры и подхода к оценке позы оригинального OpenPose.

Сеть BlazePose, разработанная Google Research и MediaPipe, использует модифицированную архитектуру BlazeFace, адаптированную для работы с полным телом, а также более легкие мобильные свёрточные нейронные сети. Эти сети оптимизированы для точной оценки позы с минимальной задержкой, что делает BlazePose популярной для приложений в фитнесе и отслеживании движений в реальном времени.

Первая версия PoseNet была выпущена в 2017 году [30]. Она построена на платформе машинного обучения TensorFlow и предоставляет реализацию HPE в реальном времени в браузере. В PoseNet встроены два алгоритма: один – для оценки позы одного человека, а другой – для оценки поз нескольких человек на входном изображении или видео. Оба алгоритма способны детектировать 17 ключевых точек на одном человеке. Время вычисления в алгоритме HPE для нескольких человек немного больше, чем для одного, но оно не зависит от количества обнаруженных людей. При использовании алгоритма для одного человека ключевые точки могут смешиваться, если на входном изображении или видео имеется более одного человека. Кроме этого, в PoseNet существуют две архитектуры – ResNet [37] и MobileNet [40]. MobileNet предназначена для мобильных устройств, она легче, чем ResNet, но имеет более низкую точность. Хотя ResNet достигает более высокой точности по сравнению с MobileNet, её большее количество слоёв требует более длительного времени загрузки и выполнения.

В 2022 году была выпущена вторая версия библиотеки PoseNet. PoseNet 2.0 была оптимизирована для меньшего потребления ресурсов и повышения производительности за счёт использования архитектуры MobileNetV2.

MoveNet (год выпуска – 2021) представляет собой модель для распознавания поз, которая в режиме реального времени определяет 17 ключевых точек на одном человеке. MoveNet существует в двух версиях: Lightning и Thunder. Точность версии Lightning ниже, чем у Thunder, но время выполнения у неё меньше.

MoveNet использует тепловые карты для точного определения местоположения ключевых точек человека. Архитектура модели состоит из двух частей: извлекателя признаков и набора предсказательных модулей.

Предсказательные модули MoveNet частично основаны на технологии CenterNet, которая применяется для повышения точности и скорости работы. CenterNet – это детектор объектов, который использует сети оценки ключевых точек для нахождения центральных точек и определения размера, местоположения и ориентации объектов.

Извлекатель признаков – это MobileNetV2 с прикреплённой сетью пирамиды признаков, которая генерирует высокоразрешённый и семантически насыщенный выходной признак.

MobileNetV2 – это нейронная сеть, оптимизированная для мобильных устройств, которая используется для извлечения признаков для обнаружения объектов, их классификации и семантической сегментации.

В предсказательных модулях MoveNet выделены четыре части: тепловая карта центра человека, поле регрессии ключевых точек, тепловая карта ключевых точек и поле 2D-смещения для каждой ключевой точки. Эти компоненты отвечают за предсказание ключевых точек человека с использованием тепловых карт.

Lightweight OpenPose была создана как упрощённая версия оригинальной OpenPose, разработанной исследовательской группой Carnegie Mellon University в 2017 году. OpenPose является одной из первых библиотек для оценки позы человека, способной отслеживать несколько ключевых точек на теле, лице и конечностях, что позволило применить её в задачах компьютерного зрения, таких как отслеживание движений, жестов и анализа поведения. Однако оригинальная OpenPose требовала значительных вычислительных ресурсов, поскольку использовала глубокую архитектуру на базе VGG-19, что делало её неподходящей для работы на мобильных устройствах и устройствах с ограниченными ресурсами.

Основное улучшение в Lightweight OpenPose достигалось за счёт использования облегченной архитектуры сети, которая сочетает компоненты MobileNet и модифицированную версию VGG. Вместо глубокой сети VGG-19, как в оригинальной версии, Lightweight OpenPose использует MobileNet в качестве базовой архитектуры для выделения признаков, что значительно снижает по-

ребление ресурсов, но сохраняет достаточную точность для большинства приложений. В процессе модификации сеть была также дополнительно сжата и оптимизирована для работы в реальном времени.

MobileNet служит основным компонентом для извлечения признаков, поскольку её свёрточные слои сконструированы так, чтобы работать быстрее и эффективнее на маломощных устройствах за счёт использования глубинно-разделимых свёрток.

Для повышения точности при минимальных затратах производительности была также применена сжимаемая архитектура с уменьшенным количеством слоев и параметров, что делает модель компактной и подходящей для встраиваемых приложений.

BlazePose была разработана в 2020 году командой Google Research и MediaPipe как специализированное решение для точной оценки человеческой позы. В отличие от предыдущих моделей для оценки позы, таких как OpenPose, основная задача BlazePose состояла в создании модели, способной обеспечивать высокие точность и производительность при минимальном потреблении ресурсов, что стало необходимым условием для мобильных приложений, особенно в фитнесе, спорте и других задачах реального времени.

BlazePose опирается на уникальную архитектуру, в основе которой лежит комбинация BlazeFace и специализированных свёрточных нейронных сетей, адаптированных для оценки позы всего тела. Перечислим ключевые архитектурные особенности.

- **Двухэтапная архитектура:** BlazePose использует два этапа для оценки позы. На первом этапе сеть обнаруживает грубую позу тела, на втором – уточняет позиции ключевых точек. Это позволяет снизить общую вычислительную нагрузку и повысить точность модели.
- **BlazeFace** как основа для BlazePose: изначально BlazeFace разрабатывалась для анализа лицевой позы в реальном времени, и её архитектура была улучшена для работы с позой тела. Основная сеть – легковесная свёрточная сеть, ориентированная на скорость и производительность на мобильных устройствах.

- **Использование MediaPipe:** BlazePose интегрирована с фреймворком MediaPipe, что позволяет использовать GPU и CPU для параллельной обработки данных и обеспечивает высокую производительность даже на устройствах со средними и низкими ресурсами. MediaPipe включает готовые инструменты для предобработки данных, оптимизированные для работы на мобильных и веб-платформах.
- **33 ключевые точки:** BlazePose выделяется в сравнении с другими сетями за счёт отслеживания 33 ключевых точек, включая детализированные точки на лице и руках, что делает её подходящей для более точного анализа движений в приложениях, таких как спортивная аналитика и отслеживание позы в фитнесе.

3. МЕТОДОЛОГИЯ ОЦЕНКИ

В этом разделе рассмотрена методология оценки работоспособности библиотек HPE, самые распространенные наборы данных и аннотаций, на которых проверяют работоспособность библиотек и их подготовку к оценке, а также ключевую метрику.

3.1. Наборы данных

Одними из самых распространенных наборов данных (датасетов), используемых в научной среде, являются Microsoft Common Object in Context (COCO), который содержит изображения и аннотации к ним, а также Penn Action, который содержит отобранные видеоролики. Характеристики этих наборов приведены в табл. 2.

Таблица 2. Характеристики библиотек оценки позы человека

Набор данных	Тип данных	Ключевых точек на человека	Предоставленная аннотация частей тела
COCO	Изображения	17	Нос, Левый глаз, Правый глаз, Левое ухо, Правое ухо, Левое Плечо, Правое плечо, Левый локоть, Правый локоть, Левое запястье, Правое запястье, Левое бедро, Правое бедро, Левое Колено, Правое Колено, Левая Лодыжка, Правая Лодыжка
Penn Action	Видеоролики	13	Голова, Левое плечо, Правое плечо, Левый локоть, Правый локоть, Левое запястье, Правое запястье, Левое бедро, Правое Бедро, Левое колено, Правое колено, Левая лодыжка, Правая лодыжка

Оба набора содержат по шесть ключевых точек для верхней и нижней частей тела. Однако есть различие в количестве аннотированных ключевых точек для головы. В COCO их пять, включая нос, глаза и уши, а в Penn Action для головы указана только одна ключевая точка.

Примеры изображений из наборов данных COCO и Penn Action можно увидеть на рис. 2 и 3 соответственно.



Рис. 2. Пример изображений из набора данных COCO

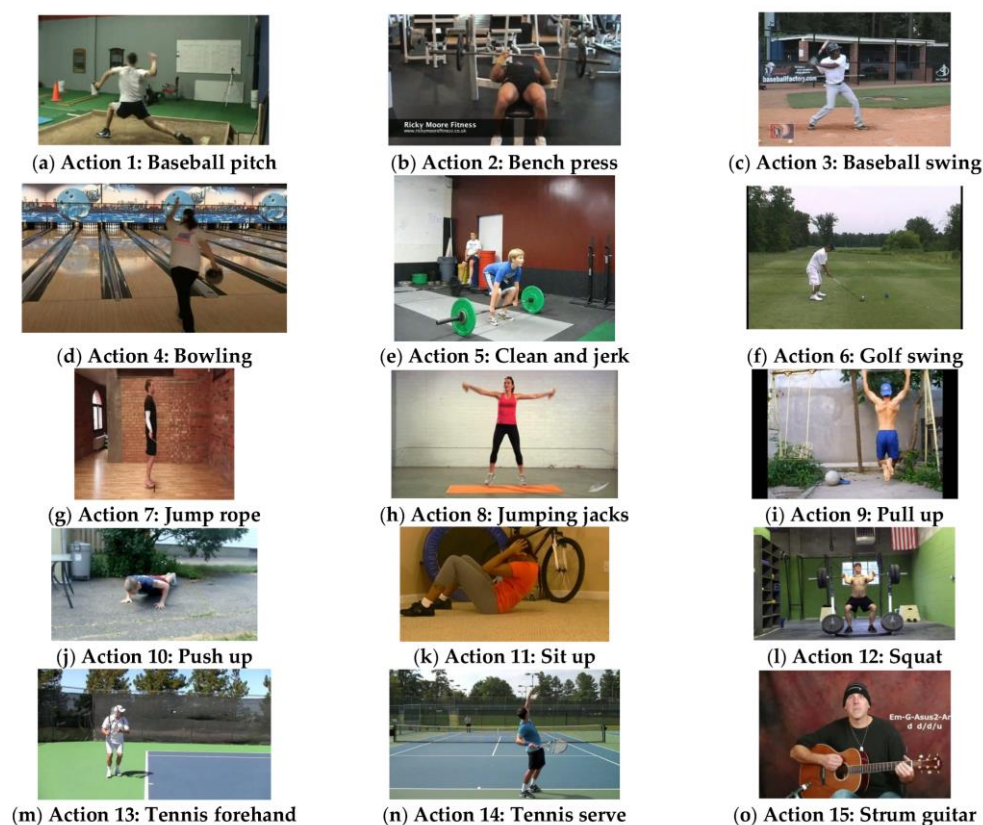


Рис. 3. Примеры кадров видео из датасета Penn Action

3.2. Подготовка наборов данных

Перед оценкой производительности библиотек НРЕ нужно провести предобработку данных для отсеивания нерелевантных данных в перечисленных выше наборах. В наборе данных COCO имеются три типа изображений: изображения с одним человеком, с несколькими людьми и без людей.

Мы рассматриваем задачу НРЕ для одного человека, поэтому изображения с несколькими людьми и без людей были удалены. Нужно также исключить изображения, на которых присутствует только часть тела. Для сравнения производительности четырех библиотек НРЕ 17 общих ключевых точек тела нужно сопоставить с 17 аннотациями, предоставленными набором данных (истинными значениями).

Что касается видео из набора Penn Action, в эксперименте использовались первые 14 действий (см. рис. 2). Поскольку Penn Action предоставил только одну

аннотацию ключевой точки головы, отличную от аннотаций в четырех библиотеках HPE (см. табл. 3), аннотация для головы была удалена из экспериментов для обеспечения объективного сравнения между библиотеками.

3.3. Метрика оценки

Метрики оценки играют важную роль в определении качества библиотек HPE. В качестве основной используется метрика «процент обнаруженных суставов» (PDJ), которая позволяет оценить производительность библиотеки HPE [37, 39]. PDJ использует евклидово расстояние между истинными и предсказанными ключевыми точками в пикселях для измерения точности обнаружения библиотек HPE. Чем выше значение PDJ, тем выше точность. Евклидово расстояние $d(x, y)$ между истинными (x_1, y_1) и предсказанными ключевыми точками (x_2, y_2) находится по формуле (1), представленной ниже.

Порог PDJ установлен на уровне 0.05 от диаметра туловища. Диаметр туловища рассчитывался как евклидово расстояние от левого плеча до правого бедра с координатами $[x_{ls}, y_{ls}]$ и $[x_{rh}, y_{rh}]$, как показано в формуле (2).

Если величина $d(x, y)$ между предсказанными и истинными ключевыми точками меньше порогового значения, предсказанная ключевая точка считалась обнаруженной корректно. Таким образом, PDJ может быть рассчитан, как показано в формуле (3), где n обозначает общее количество предсказанных суставов.

$$d(x, y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \text{ (pixel)} \quad (1)$$

$$\text{torso diameter} = \sqrt{(x_{ls} - x_{rh})^2 + (y_{ls} - y_{rh})^2} \quad (2)$$

$$PDJ = \frac{\sum_{i=1}^n \text{bool}(d_i < 0.05 \times \text{torso diameter})}{n} \quad (3)$$

4. Преимущества и недостатки рассматриваемых библиотек

В этом разделе рассмотрены четыре популярных библиотеки для НРЕ: PoseNet, OpenPose, MoveNet и BlazePose. Представлены ключевые преимущества и недостатки каждой из них, установленные на основе анализа научных работ и открытых исходных кодов.

4.1. PoseNet

Библиотека PoseNet и её версии на основе TensorFlow.js отличаются простотой интеграции и использования, что делает её идеальным инструментом для разработки практических приложений в браузере без необходимости установки сложных зависимостей [40].

Модель PoseNet оптимизирована для работы с мобильными и веб-приложениями. Её цель – обеспечить работу в режиме реального времени на устройствах с ограниченными ресурсами, что упрощает интеграцию в пользовательские сервисы.

Кроме того, PoseNet не требует специальных знаний и навыков, что делает её доступной широкому кругу пользователей. Благодаря подробной документации и готовым примерам модель часто используется в образовательных и любительских проектах. Но при этом библиотека PoseNet создана для быстрого и лёгкого определения ключевых точек, поэтому точность может быть ниже, чем у более сложных моделей [40].

PoseNet ориентирована на работу с одним объектом в кадре. Для распознавания нескольких объектов одновременно требуется дополнительная настройка или использование других методов.

4.2. OpenPose

Это одна из первых библиотек, которая позволяет определять позы нескольких человек в реальном времени с высоким качеством.

OpenPose использует карту связей (Part Affinity Fields), что позволяет эффективно устанавливать взаимосвязи между конечностями и сегментировать несколько человек даже при наличии перекрытий [41].

OpenPose широко известна и активно используется в научных исследованиях и промышленных прототипах. Но при этом OpenPose требует значительных

ресурсов. Модель занимает много памяти и нагружает процессор и графический процессор, что затрудняет её использование на мобильных устройствах и в приложениях с ограниченными вычислительными ресурсами [41].

OpenPose требует более глубокого понимания параметров и окружения для корректного развёртывания, особенно при необходимости ускорения вычислений. В отличие от PoseNet и MoveNet, OpenPose не так проста в настройке.

4.3. MoveNet

Библиотека MoveNet разработана с учётом ограниченных вычислительных ресурсов и предназначена для работы в режиме реального времени, включая мобильные устройства [42]. Это позволяет использовать ее в различных сферах, в том числе в медицине, спорте и робототехнике.

Существует несколько версий MoveNet, которые позволяют выбирать между точностью и скоростью. Например, версии Thunder и Lightning отличаются по этим параметрам.

MoveNet легко интегрируется с экосистемой TensorFlow. Модель доступна через TensorFlow Hub и может быть легко встроена в различные приложения, поддерживающие TF Lite, что упрощает разработку и развёртывание этой библиотеки. MoveNet хорошо подходит для отслеживания общих положений тела, но ее архитектура не всегда обеспечивает такую же точность локализации мелких деталей, как у более тяжёлых моделей. MoveNet использует упрощённый набор ключевых точек, аналогичный COCO, что может быть недостаточно для некоторых специализированных задач [42], например для анализа движений в танце или в медицинских целях.

4.4. BlazePose

Библиотека BlazePose, разработанная в рамках MediaPipe, способна отслеживать до 33 ключевых точек, включая дополнительные ориентиры кистей и стоп, что обеспечивает более детальное представление о позе.

Перечислим основные преимущества этой библиотеки.

1) Оптимизация для мобильных устройств: BlazePose, как и другие решения MediaPipe, предназначены для работы в режиме реального времени

на смартфонах и планшетах, поддерживая ускорение вычислений на графическом процессоре [44].

2) Гибкость в сочетании с другими моделями: благодаря интеграции в MediaPipe, BlazePose часто используется совместно с другими библиотеками для распознавания жестов и выражений лица, формируя комплексные решения в области компьютерного зрения [45, 46].

3) Зависимость от инфраструктуры MediaPipe: для разработки с использованием BlazePose необходимо понимание экосистемы MediaPipe, что может усложнить интеграцию в сторонние проекты.

Таким образом, каждая из рассмотренных библиотек обладает собственным набором преимуществ и ограничений, что позволяет выбирать конкретное решение под определённые задачи – от мобильных приложений и веб-сервисов (PoseNet, MoveNet, BlazePose) до комплексных систем отслеживания позы множества людей в реальном времени (OpenPose).

ЗАКЛЮЧЕНИЕ

Рассмотрев самые распространённые библиотеки определения позы человека (HPE), а также метрику оценки их качества, можно сделать вывод, что для решения задачи определения позы человека оптимальным выбором будет библиотека BlazePose, так как она способна отслеживать 33 ключевые точки и при этом показывает лучший результат в условиях работы на мобильных устройствах. В будущем планируется представить результаты эксперимента детекции позы человека, подтверждающие выводы, сформулированные выше, а в качестве ключевой метрики использовать предложенную выше метрику, равную проценту правильно определенных суставов.

СПИСОК ЛИТЕРАТУРЫ

1. *Su C., Li J.; Zhang S., Xing J., Gao W., Tian Q.* Pose-driven deep convolutional model for person re-identification // *Proceedings of the IEEE International Conference on Computer Vision*. 2017. P. 3960–3969.
2. *Xu J., Zhao R., Zhu F., Wang H., Ouyang W.* Attention-aware compositional network for person re-identification // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018. P. 2119–2128.

3. *Thyagarajmurthy A., Ninad M.G., Rakesh B.G., Niranjan S., Manvi B.* Anomaly detection in surveillance video using pose estimation // *Proceedings of the Emerging Research in Electronics, Computer Science and Technology*; Springer. 2019. P. 753–766.
4. *Lamas A., Tabik S., Montes A.C., Pérez-Hernández F., García J., Olmos R., Herrera F.* Human pose estimation for mitigating false negatives in weapon detection in video-surveillance // *Neurocomputing*. 2022. Vol. 489. P. 488–503.
5. *Yoo H.R., Lee B.H.* An OpenPose-based child abuse decision system using surveillance video // *J. Korea Inst. Inf. Commun. Eng.* 2019. Vol. 23. P. 282–290.
6. *Park J.H., Song K., Kim Y.-S.* A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems // *J. Korea Soc. Comput. Inf.* 2018. Vol. 23. P. 9–16.
7. *Chang Y.J., Chen S.F., Huang J.D.* A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities // *Res. Dev. Disabil.* 2011. Vol. 32. P. 2566–2570.
8. *Hassan H.A., Abdallah B.H., Abdallah A.A., Abdel-Aal R.O., Numan R.R., Darwish A.K., El-Behaidy W.H.* Automatic Feedback for Physiotherapy Exercises Based On PoseNet // *FCAI-Inform. Bull.* 2020. Vol. 2. P. 10–14.
9. *Shapoval S., García Zapirain B., Mendez Zorrilla A., Mugueta-Aguinaga I.* Biofeedback applied to interactive serious games to monitor frailty in an elderly population // *Appl. Sci.* 2021. Vol. 11. P. 3502.
10. *Chua J., Ong L.Y., Leow M.C.* Telehealth using PoseNet-based system for in-home rehabilitation // *Future Internet*. 2021. Vol. 13. P. 173.
11. *Kim W., Sung J., Saakes D., Huang C., Xiong S.* Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose) // *Int. J. Ind. Ergon.* 2021. Vol. 84. P. 103164.
12. *Jawale C.D., Joshi K.A., Gogate S.K., Badgujar C.* Elcare: Elderly Care with Fall Detection // *J. Phys. Conf. Ser.* 2022. Vol. 2273. P. 012019.
13. *Kapoor R., Jaiswal A., Makedon F.* Light-weight seated posture guidance

system with machine learning and computer vision // Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments. 2022. P. 595–600.

14. Landi H. Google, ProMedica Team up with Include Health to Tap into Growing Virtual MSK Market // Fierce Healthcare. URL: <https://www.fiercehealthcare.com/tech/google-promedica-team-up-include-health-to-tap-into-virtual-msk-market>

15. Chen W., Jiang Z., Guo H., Ni X. Fall detection based on key points of human skeleton using OpenPose // Symmetry. 2020. Vol. 12. P. 744.

16. Zou J., Li B., Wang L., Li Y., Li X., Lei R., Sun S. Intelligent fitness trainer system based on human pose estimation // Proceedings of the International Conference On Signal and Information Processing, Networking and Computers. 2018. P. 593–599.

17. Suda S., Makino Y., Shinoda H. Prediction of volleyball trajectory using skeletal motions of setter player // Proceedings of the 10th Augmented Human International Conference. 2019. P. 1–8.

18. Wang J., Qiu K., Peng H., Fu J., Zhu J. AI coach: Deep human pose estimation and analysis for personalized athletic training assistance // Proceedings of the 27th ACM International Conference on Multimedia. 2019. P. 374–382.

19. Jeon H., Yoon Y., Kim D. Lightweight 2D human pose estimation for fitness coaching system // Proceedings of the 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC). 2021. P. 1–4.

20. Li Y.C., Chang C.T., Cheng C.C., Huang Y.L. Baseball Swing Pose Estimation Using OpenPose // Proceedings of the 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI). 2021. P. 6–9.

21. Park H.J., Baek J.W., Kim J.H. Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose // Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE). 2020. P. 1389–1394.

22. Nguyen H.T.P., Woo Y., Huynh N.N., Jeong H. Scoring of Human Body-Balance Ability on Wobble Board Based on the Geometric Solution // Appl. Sci. 2022. Vol. 12. P. 5967.

23. *Patil A., Rao D., Utturwar K., Shelke T., Sarda E.* Body Posture Detection and Motion Tracking using AI for Medical Exercises and Recommendation System // ITM Web Conf. 2022. Vol. 44. P. 03043.

24. *Devanandan M., Rasaratnam V., Anbalagan M.K., Asokan N., Panchendrarajan R., Tharmaseelan J.* Cricket Shot Image Classification Using Random Forest // Proceedings of the 2021 3rd International Conference on Advancements in Computing (ICAC). 2021. P. 425–430.

25. *Joseph R., Ayyappan M., Shetty T., Gaonkar G., Nagpal A.* BeFit—A Real-Time Workout Analyzer // Proceedings of the Sentimental Analysis and Deep Learning. 2022. P. 303–318.

26. *Mahendran N.* Deep Learning for Fitness // arXiv. 2021. arXiv:2109.01376.

27. *Agarwal S., Gupta M., Khandelwal S., Jain P., Aggarwal A., Singh D., Mishra V.K.* FitMe: A Fitness Application for Accurate Pose Estimation Using Deep Learning // Proceedings of the 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). 2021. P. 232–237.

28. *Nakai M., Tsunoda Y., Hayashi H., Murakoshi H.* Prediction of basketball free throw shooting by OpenPose // Proceedings of the JSAI International Symposium on Artificial Intelligence. 2018. P. 435–446.

29. *Zheng C., Wu W., Chen C., Yang T., Zhu S., Shen J., Kehtarnavaz N., Shah M.* Deep learning-based human pose estimation: A survey // arXiv. 2020, arXiv:2012.13392.

30. *Papandreou G., Zhu T., Kanazawa N., Toshev A., Thompson J., Bregler C., Murphy K.* Towards accurate multi-person pose estimation in the wild // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu. 2017. P. 4903–4911.

31. *Jo B., Kim S.* Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices // Trait. du Signal. 2022. Vol. 39. P. 119–124.

32. *Cao Z., Simon T., Wei S.E., Sheikh Y.* Realtime multi-person 2d pose estima-

tion using part affinity fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 7291–7299.

33. *Erramchetty S.K. et al.* Human pose estimation using blaze pose // AIP Conference Proceedings. 2024. Vol. 2971. No. 1.

34. *Gadhiya R., Kalani N.* Analysis of Deep Learning Based Pose Estimation Techniques for Locating Landmarks on Human Body Parts // Proceedings of the 2021 International Conference on Circuits, Controls and Communications (CCUBE). 2021. P. 1–4.

35. *Lin T.Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick C.L.* Microsoft coco: Common objects in context // Proceedings of the European Conference on Computer Vision. 2014. P. 740–755.

36. *Andriluka M., Pishchulin L., Gehler P., Schiele B.* 2D human pose estimation: New benchmark and state of the art analysis // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 3686–3693.

37. *Toshev A., Szegedy C.* DeepPose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1653–1660.

38. *Chung J.-L., Ong L.-Y., Leow M.-C.* Comparative Analysis of Skeleton-Based Human Pose Estimation // Future Internet. 2022. Vol. 14, No. 12. P. 380.
<https://doi.org/10.3390/fi14120380>

39. *Ahmedt-Aristizabal D., Nguyen K., Denman S., Sridharan S., Dionisio S., Fookes C.* Deep motion analysis for epileptic seizure classification // Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018. P. 3578–3581.

40. *Ishaq A., Haq I.U.* Real-time single-person pose estimation on low-power embedded devices // 2019 International Conference on Frontiers of Information Technology (FIT).

41. *Cao Z., Simon T., Wei S.-E., Sheikh Y.* Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 7291–7299.

42. Papandreou G., Zhu T., Chen L.-C., Gidaris S., Thompson J., Murphy K. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model // Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 269–286.

43. TensorFlow Blog. MoveNet: Ultra fast and accurate pose detection model. 2021.

URL: <https://blog.tensorflow.org/2021/05/movenet-and-posture-analysis.html>

44. Bazarevsky V., Zhang F. BlazePose: On-Device Real-Time Body Pose Tracking. Google AI Blog. 2020.

URL: <https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html>

45. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for real-time image recognition through augmented reality devices // The 2023 International Conference on Artificial Life and Robotics (ICAROB2023). 2023. P. 868–874.

46. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for augmented reality applications // Journal of Advances in Artificial Life Robotics. 2023. Vol. 3. Iss. 3(3). P. 179–184.

47. Бакиров А.Р. и др. Использование контроллера Microsoft Kinect в разработке реабилитационных игр // Электронные библиотеки. 2016. Т. 19. №. 6. С. 521–537.

48. You Y., Zhao Y. A human pose estimation algorithm based on the integration of improved convolutional neural networks and multi-level graph structure constrained model // Pers. Ubiquitous Comput. 2019. Vol. 23. P. 607–616.

COMPARATIVE ANALYSIS OF LIBRARIES FOR HUMAN POSE DETECTION IN MOBILE DEVICE ENVIRONMENTS

E. I. Yarko^[0009-0009-0502-1066]

Kazan Federal university, Kazan, 420008, Russia

yarkoeg@gmail.com

Abstract

Human Pose Estimation (HPE) has become one of the most relevant topics in computer vision research. This technology can be applied in various fields such as video surveillance, medical care, and sports motion analysis. Due to the increasing demand for HPE, many libraries for this technology have been developed in the last 20 years. Since 2017, many HPE algorithms based on skeletal model have been published and packaged into libraries for easy use by researchers.

These libraries are important for researchers who want to integrate them into real-world applications for video surveillance, medical care, and sports motion analysis.

This paper investigates the strengths and weaknesses of four popular HPE advanced human pose recognition libraries that can run on mobile devices: Lightweight OpenPose, PoseNet, MoveNet, and Blase Pose.

Keywords: *Human pose, Human Pose Estimation, HPE, pose detection, computer vision, mobile devices, augmented reality, Lightweight OpenPose, PoseNet, MoveNet, BlazePose, skeletal model.*

REFERENCES

1. Su C., Li J.; Zhang S., Xing J., Gao W., Tian Q. Pose-driven deep convolutional model for person re-identification // Proceedings of the IEEE International Conference on Computer Vision. 2017. P. 3960–3969.
2. Xu J., Zhao R., Zhu F., Wang H., Ouyang W. Attention-aware compositional network for person re-identification // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018. P. 2119–2128.
3. Thyagarajmurthy A., Ninad M.G., Rakesh B.G., Niranjana S., Manvi B.

Anomaly detection in surveillance video using pose estimation // Proceedings of the Emerging Research in Electronics, Computer Science and Technology; Springer. 2019. P. 753–766.

4. *Lamas A., Tabik S., Montes A.C., Pérez-Hernández F., García J., Olmos R., Herrera F.* Human pose estimation for mitigating false negatives in weapon detection in video-surveillance // Neurocomputing. 2022. Vol. 489. P. 488–503.

5. *Yoo H.R., Lee B.H.* An OpenPose-based child abuse decision system using surveillance video // J. Korea Inst. Inf. Commun. Eng. 2019. Vol. 23. P. 282–290.

6. *Park J.H., Song K., Kim Y.-S.* A Kidnapping Detection Using Human Pose Estimation in Intelligent Video Surveillance Systems // J. Korea Soc. Comput. Inf. 2018. Vol. 23. P. 9–16.

7. *Chang Y.J., Chen S.F., Huang J.D.* A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities // Res. Dev. Disabil. 2011. Vol. 32. P. 2566–2570.

8. *Hassan H.A., Abdallah B.H., Abdallah A.A., Abdel-Aal R.O., Numan R.R., Darwish A.K., El-Behaidy W.H.* Automatic Feedback for Physiotherapy Exercises Based On PoseNet // FCAI-Inform. Bull. 2020. Vol. 2. P. 10–14.

9. *Shapoval S., García Zapirain B., Mendez Zorrilla A., Mugueta-Aguinaga I.* Biofeedback applied to interactive serious games to monitor frailty in an elderly population // Appl. Sci. 2021. Vol. 11. P. 3502.

10. *Chua J., Ong L.Y., Leow M.C.* Telehealth using PoseNet-based system for in-home rehabilitation // Future Internet. 2021. Vol. 13. P. 173.

11. *Kim W., Sung J., Saakes D., Huang C., Xiong S.* Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose) // Int. J. Ind. Ergon. 2021. Vol. 84. P. 103164.

12. *Jawale C.D., Joshi K.A., Gogate S.K., Badgujar C.* Elcare: Elderly Care with Fall Detection // J. Phys. Conf. Ser. 2022. Vol. 2273. P. 012019.

13. *Kapoor R., Jaiswal A., Makedon F.* Light-weight seated posture guidance system with machine learning and computer vision // Proceedings of the 15th International Conference on Pervasive Technologies Related to Assistive Environments.

2022. P. 595–600.

14. *Landi H.* Google, ProMedica Team up with Include Health to Tap into Growing Virtual MSK Market // *Fierce Healthcare*.

URL: <https://www.fiercehealthcare.com/tech/google-promedica-team-up-include-health-to-tap-into-virtual-msk-market>

15. *Chen W., Jiang Z., Guo H., Ni X.* Fall detection based on key points of human-skeleton using OpenPose // *Symmetry*. 2020. Vol. 12. P. 744.

16. *Zou J., Li B., Wang L., Li Y., Li X., Lei R., Sun S.* Intelligent fitness trainer system based on human pose estimation // *Proceedings of the International Conference On Signal and Information Processing, Networking and Computers*. 2018. P. 593–599.

17. *Suda S., Makino Y., Shinoda H.* Prediction of volleyball trajectory using skeletal motions of setter player // *Proceedings of the 10th Augmented Human International Conference*. 2019. P. 1–8.

18. *Wang J., Qiu K., Peng H., Fu J., Zhu J.* AI coach: Deep human pose estimation and analysis for personalized athletic training assistance // *Proceedings of the 27th ACM International Conference on Multimedia*. 2019. P. 374–382.

19. *Jeon H., Yoon Y., Kim D.* Lightweight 2D human pose estimation for fitness coaching system // *Proceedings of the 2021 36th International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC)*. 2021. P. 1–4.

20. *Li Y.C., Chang C.T., Cheng C.C., Huang Y.L.* Baseball Swing Pose Estimation Using OpenPose // *Proceedings of the 2021 IEEE International Conference on Robotics, Automation and Artificial Intelligence (RAAI)*. 2021. P. 6–9.

21. *Park H.J., Baek J.W., Kim J.H.* Imagery based Parametric Classification of Correct and Incorrect Motion for Push-up Counter Using OpenPose // *Proceedings of the 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*. 2020. P. 1389–1394.

22. *Nguyen H.T.P., Woo Y., Huynh N.N., Jeong H.* Scoring of Human Body-Balance Ability on Wobble Board Based on the Geometric Solution // *Appl. Sci*. 2022. Vol. 12. P. 5967.

23. *Patil A., Rao D., Utturwar K., Shelke T., Sarda E.* Body Posture Detection and Motion Tracking using AI for Medical Exercises and Recommendation System // ITM Web Conf. 2022. Vol. 44. P. 03043.
24. *Devanandan M., Rasaratnam V., Anbalagan M.K., Asokan N., Panchendrarajan R., Tharmaseelan J.* Cricket Shot Image Classification Using Random Forest // Proceedings of the 2021 3rd International Conference on Advancements in Computing (ICAC). 2021. P. 425–430.
25. *Joseph R., Ayyappan M., Shetty T., Gaonkar G., Nagpal A.* BeFit—A Real-Time Workout Analyzer // Proceedings of the Sentimental Analysis and Deep Learning. 2022. P. 303–318.
26. *Mahendran N.* Deep Learning for Fitness // arXiv. 2021. arXiv:2109.01376.
27. *Agarwal S., Gupta M., Khandelwal S., Jain P., Aggarwal A., Singh D., Mishra V.K.* FitMe: A Fitness Application for Accurate Pose Estimation Using Deep Learning // Proceedings of the 2021 2nd International Conference on Secure Cyber Computing and Communications (ICSCCC). 2021. P. 232–237.
28. *Nakai M., Tsunoda Y., Hayashi H., Murakoshi H.* Prediction of basketball free throw shooting by OpenPose // Proceedings of the JSAI International Symposium on Artificial Intelligence. 2018. P. 435–446.
29. *Zheng C., Wu W., Chen C., Yang T., Zhu S., Shen J., Kehtarnavaz N., Shah M.* Deep learning-based human pose estimation: A survey // arXiv. 2020, arXiv:2012.13392.
30. *Papandreou G., Zhu T., Kanazawa N., Toshev A., Tompson J., Bregler C., Murphy K.* Towards accurate multi-person pose estimation in the wild // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu. 2017. P. 4903–4911.
31. *Jo B., Kim S.* Comparative Analysis of OpenPose, PoseNet, and MoveNet Models for Pose Estimation in Mobile Devices // Trait. du Signal. 2022. Vol. 39. P. 119–124.

32. Cao Z., Simon T., Wei S.E., Sheikh Y. Realtime multi-person 2d pose estimation using part affinity fields // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017. P. 7291–7299.
33. Erramchetty S.K. et al. Human pose estimation using blaze pose // AIP Conference Proceedings. 2024. Vol. 2971. No. 1.
34. Gadhiya R., Kalani N. Analysis of Deep Learning Based Pose Estimation Techniques for Locating Landmarks on Human Body Parts // Proceedings of the 2021 International Conference on Circuits, Controls and Communications (CCUBE). 2021. P. 1–4.
35. Lin T.Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollár P., Zitnick C.L. Microsoft coco: Common objects in context // Proceedings of the European Conference on Computer Vision. 2014. P. 740–755.
36. Andriluka M., Pishchulin L., Gehler P., Schiele B. 2D human pose estimation: New benchmark and state of the art analysis // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 3686–3693.
37. Toshev A., Szegedy C. DeepPose: Human pose estimation via deep neural networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1653–1660.
38. Chung J.-L., Ong L.-Y., Leow M.-C. Comparative Analysis of Skeleton-Based Human Pose Estimation // Future Internet. 2022. Vol. 14, No. 12. P. 380.
<https://doi.org/10.3390/fi14120380>
39. Ahmedt-Aristizabal D., Nguyen K., Denman S., Sridharan S., Dionisio S., Fookes C. Deep motion analysis for epileptic seizure classification // Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2018. P. 3578–3581.
40. Ishaq A., Haq I.U. Real-time single-person pose estimation on low-power embedded devices // 2019 International Conference on Frontiers of Information Technology (FIT).
41. Cao Z., Simon T., Wei S.-E., Sheikh Y. Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields // Proceedings of the IEEE Conference on Computer

Vision and Pattern Recognition (CVPR). 2017. P. 7291–7299.

42. Papandreou G., Zhu T., Chen L.-C., Gidaris S., Tompson J., Murphy K. PersonLab: Person Pose Estimation and Instance Segmentation with a Bottom-Up, Part-Based, Geometric Embedding Model // Proceedings of the European Conference on Computer Vision (ECCV). 2018. P. 269–286.

43. TensorFlow Blog. MoveNet: Ultra fast and accurate pose detection model. 2021.

URL: <https://blog.tensorflow.org/2021/05/movenet-and-posture-analysis.html>

44. Bazarevsky V., Zhang F. BlazePose: On-Device Real-Time Body Pose Tracking. Google AI Blog. 2020.

URL: <https://ai.googleblog.com/2020/08/on-device-real-time-body-pose-tracking.html>

45. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for real-time image recognition through augmented reality devices // The 2023 International Conference on Artificial Life and Robotics (ICAROB2023). 2023. P. 868–874.

46. Gabdullina D.R., Zykov E.Yu., Kugurakova V.V. Using OpenCV for augmented reality applications // Journal of Advances in Artificial Life Robotics. 2023. Vol. 3. Iss. 3(3). P. 179–184.

47. Bakirov A.R. et al. Using Microsoft Kinect in the rehabilitation games development // Russian Digital Libraries Journal. 2016. V. 19. No. P. 521–537.

48. You Y., Zhao Y. A human pose estimation algorithm based on the integration of improved convolutional neural networks and multi-level graph structure constrained model // Pers. Ubiquitous Comput. 2019. Vol. 23. P. 607–616.

СВЕДЕНИЯ ОБ АВТОРЕ



ЯРКО Егор Игоревич – аспирант кафедры индустрии разработки игр института ИТИС. Сфера научных интересов – приложения дополненной реальности с детектированием позы человека в реальном времени.

Egor Igorevich YARKO – postgraduate student of the Chair of Game Development Industry at ITIS Institute. His research interests include augmented reality applications with real-time human pose detection.

email: yarkoeg@gmail.com

ORCID: 0009-0009-0502-1066

Материал поступил в редакцию 28 января 2025 года

ИНФОРМАЦИОННО-АНАЛИТИЧЕСКАЯ СИСТЕМА СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ НЕЙРО-НЕЧЕТКОГО ПОДХОДА

М. В. Бобырь¹ [0000-0002-5400-6817], Б. А. Бондаренко² [0000-0001-5415-9015]

^{1, 2}Юго-Западный государственный университет, Курск, 305040, Россия

¹maxbobyry@gmail.com, ²sikersinko@gmail.com

Аннотация

Представлена информационно-аналитическая система (ИАС) для высокоскоростной сегментации изображений в градациях серого, основанной на модифицированном методе дефаззификации с использованием треугольных функций принадлежности. Цель исследования заключается в анализе влияния упрощения формулы дефаззификации на точность и контрастность выделения объектов. Предложенный подход включает адаптивное обучение весового коэффициента, позволяющее динамически корректировать процесс дефаззификации в зависимости от целевых значений. Проведено сравнение базового метода усреднения значений принадлежности и модифицированного варианта с учетом нелинейных весов. Эксперименты, проведенные на изображениях формата 1024x720, продемонстрировали, что разработанная ИАС обеспечивает высокую точность сегментации и улучшенную контрастность объектов при минимальных вычислительных затратах. Результаты подтверждают превосходство предложенного метода над традиционными подходами, подчеркивая перспективы применения искусственного интеллекта в задачах компьютерного зрения.

Ключевые слова: ИАС, нейро-нечеткий алгоритм, сегментация изображений, дефаззификация, искусственный интеллект, метод отношения площадей.

ВВЕДЕНИЕ

В последние годы обработка изображений стала одной из ключевых областей применения искусственного интеллекта и компьютерного зрения. Особое внимание уделяется методам сегментации изображений, которые позволяют выделять значимые объекты или области на основе их характеристик [1, 2]. Одним

из популярных подходов к решению задачи обработки изображений является использование нечеткой логики, которая эффективно справляется с неопределенностью и переходными состояниями в данных. Центральным этапом в таких системах является дефаззификация – процесс преобразования нечетких значений принадлежности в четкие выходные результаты. Традиционные методы дефаззификации, такие как вычисление центра тяжести или среднего значения, часто требуют компромисса между точностью и вычислительной сложностью [3].

В настоящей работе рассмотрено влияние упрощения формулы дефаззификации на результаты сегментации изображений с использованием треугольных функций принадлежности [4–6]. Проведено сравнение двух подходов: базового метода, основанного на усреднении значений принадлежности, и модифицированного варианта с учетом нелинейных весов. Кроме того, предложен алгоритм с адаптивным обучением весового коэффициента, который позволяет динамически корректировать процесс дефаззификации в зависимости от целевых значений. На основе экспериментов с реальными изображениями проведен анализ, как такие изменения сказываются на контрастности и точности выделения объектов, а также оценена практическая применимость разработанных программных решений. Результаты исследования будут полезны для дальнейшего совершенствования методов нечеткой сегментации в задачах компьютерного зрения.

В настоящей работе представлена модель нейро-нечеткого алгоритма для высокоскоростной сегментации изображений в градации серого. Эксперименты на изображениях формата 1024x720 показали эффективность предлагаемого метода. Реализованный искусственный интеллект (ИИ) подчёркивает потенциал применения ИИ в задачах сегментации изображений с минимальными вычислительными затратами [7].

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ НЕЙРО-НЕЧЕТКОГО АЛГОРИТМА СЕГМЕНТАЦИИ

Для вычисления результирующей переменной в алгоритме сегментации используем нейро-нечеткую модель обучения с дефаззификатором на основе модифицированного метода отношения площадей [8]. Рассмотрим этапы построения этого метода дефаззификации.

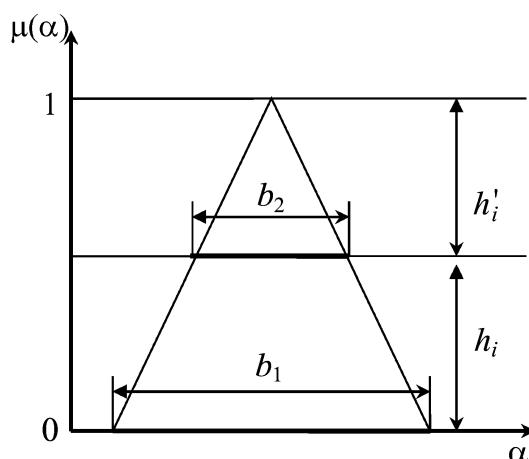


Рис. 1. Упрощение процесса вычисления верхнего основания усеченной функции принадлежности

Шаг 1. Определим общую площадь фигуры $S_{общ}$, описывающую выходную нечеткую переменную (рис. 2, а):

$$S_{общ} = \frac{nb_1\omega_i}{2}, \quad (1)$$

где n – число термов выходной нечеткой переменной ($n = 5$), b_1 – длина нижнего основания выходных ФП, ω_i – весовой коэффициент, i – номер термина ($i = 1, \dots, n$).

Шаг 2. Проведем усечение термов выходной ФП (см. рис. 2, б):

$$\mu(\alpha)' = \min_{i=1}^n [\mu(\alpha_i); h_i], \quad (2)$$

где $\mu(\alpha_i)$ – термы выходной ФП (см. рис. 2, а), $\min$ – мягкая операция нечеткого минимума, h_i – высота усеченного термина (см. рис. 1).

Шаг 3. Найдем площадь каждого из усеченных термов:

$$S_i = \begin{cases} 0, & \text{если } h_i = 0; \\ \frac{b_1}{2}, & \text{если } h_i = 1; \\ \frac{h_i}{2}(b_1 + b_2), & \text{если } h_i \in (0,1), \end{cases} \quad (3a)$$

$$(3b)$$

$$(3b)$$

где b_2 – длина верхнего основания усеченных термов выходной нечеткой переменной. Параметры треугольной функций принадлежности (ФП) приведены на рис. 1.

Шаг 4. Рассчитаем общую суммарную площадь $S_{\text{общ.сум.пл}}$ усеченных термов выходной нечеткой переменной:

$$S_{\text{общ.сум.пл}} = \sum_{i=1}^n S_i. \quad (4)$$

Шаг 5. Вычислим отношение площади усеченных термов к общей площади:

$$D = \frac{S_{\text{общ.сум.пл}}}{S_{\text{общ}}}. \quad (5)$$

Шаг 6. Определим дефаззификацию четкого значения выходной переменной на основе линейной функции:

$$\text{MAR} = [D(\alpha_K - \alpha_H)] + \alpha_H, \quad (6)$$

где α_H , α_K – соответственно начальное и конечное значения суппорта выходной переменной ($\alpha_H=0,4$; $\alpha_K=1$, см. рис. 2, а).

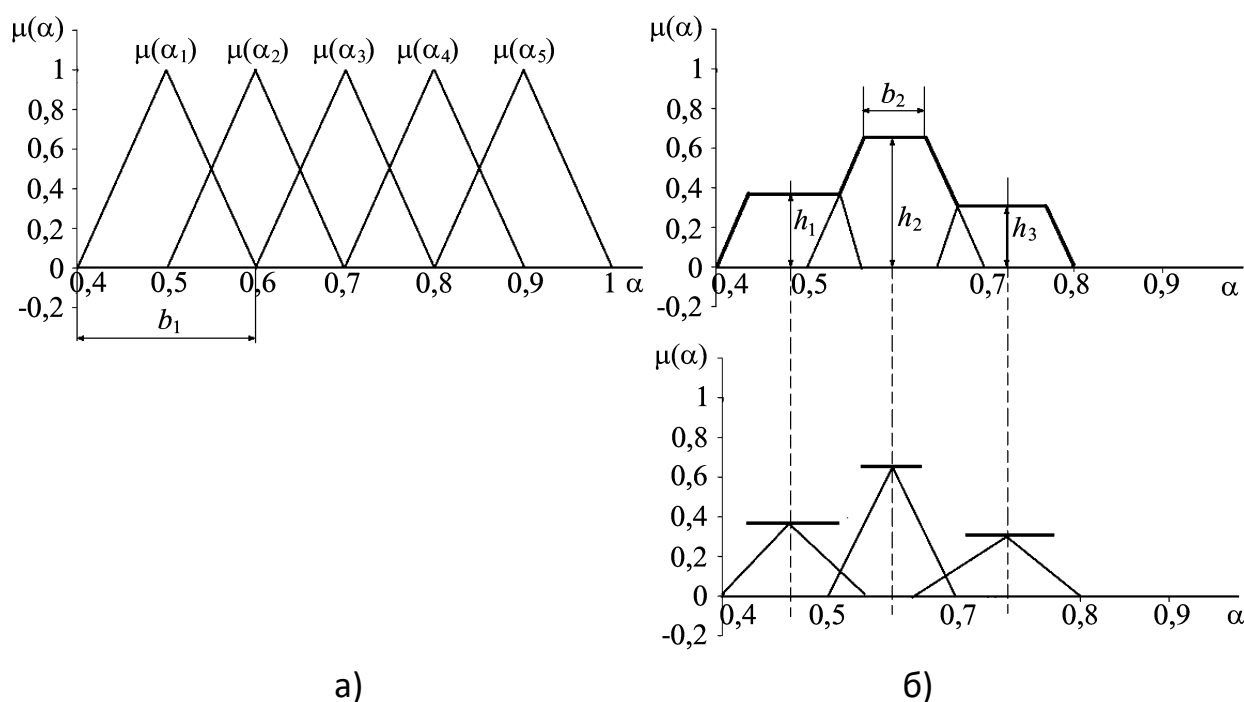


Рис. 2. Выходная нечеткая переменная: а – функции принадлежности; б – усеченные термы

Путем экспериментальных исследований метода отношения площадей установлено, что формула (3в) – самая трудоемкая операция. Поэтому проведем следующие модификации.

Шаг 7. Предположим, что усеченные термы выходной переменной после агрегации имеют вид треугольной функции и не зависят от длины верхнего основания усеченного терма b_2 (см. рис. 2, б, нижняя проекция). Тогда формулу (3в) сведем к виду

$$S_i = \frac{h_i b_1}{2}. \quad (7)$$

Шаг 8. Подставим формулу (7) в уравнение (4) и получим выражение для нахождения суммарной площади $S_{\text{общ. сум. пл}}$ усеченных термов выходной нечеткой переменной:

$$S_{\text{общ. сум. пл}} = \frac{\sum_{i=1}^n h_i b_1}{2}. \quad (8)$$

Шаг 9. Подставим формулы (1) и (8) в выражение (5) и получим формулу расчета отношения общей площади к суммарной площади усеченных термов:

$$D = \frac{\sum_{i=1}^n h_i}{n \omega_i}. \quad (9)$$

Шаг 10. Подставим уравнение (9) в формулу (6) и получим уравнение для нахождения четкой величины выходной переменной:

$$\text{MAR2} = \left[\frac{\sum_{i=1}^n h_i}{n \omega_i} (\alpha_{\kappa} - \alpha_{\mu}) \right] + \alpha_{\mu} \quad (10)$$

Шаг 11. После процесса дефаззификации выполняется обучение системы. Выходное значение после обучения определяется по формуле (11):

$$\begin{aligned} \omega_{i+1} &= \omega_i + ([\text{MAR2} - \alpha_{\text{target}}] \times \text{Mult}), \\ &\text{until } |\text{MAR2} - \alpha_{\text{target}}| \leq T, \end{aligned} \quad (11)$$

где Mult – скорость обучения (по умолчанию 0.04), T – пороговый коэффициент (по умолчанию 0.01), α_{target} – ожидаемое выходное значение после дефаззификации. Таким образом, выполняется i итераций, пока значение ω_{i+1} не достигнет целевого значения.

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ

Онтологическая модель нейро-нечеткой сегментации изображений с функцией обучения (рис. 2) предоставляет собой ИАС, позволяющую сегментировать объект и фон на изображении посредством нечеткой логики.

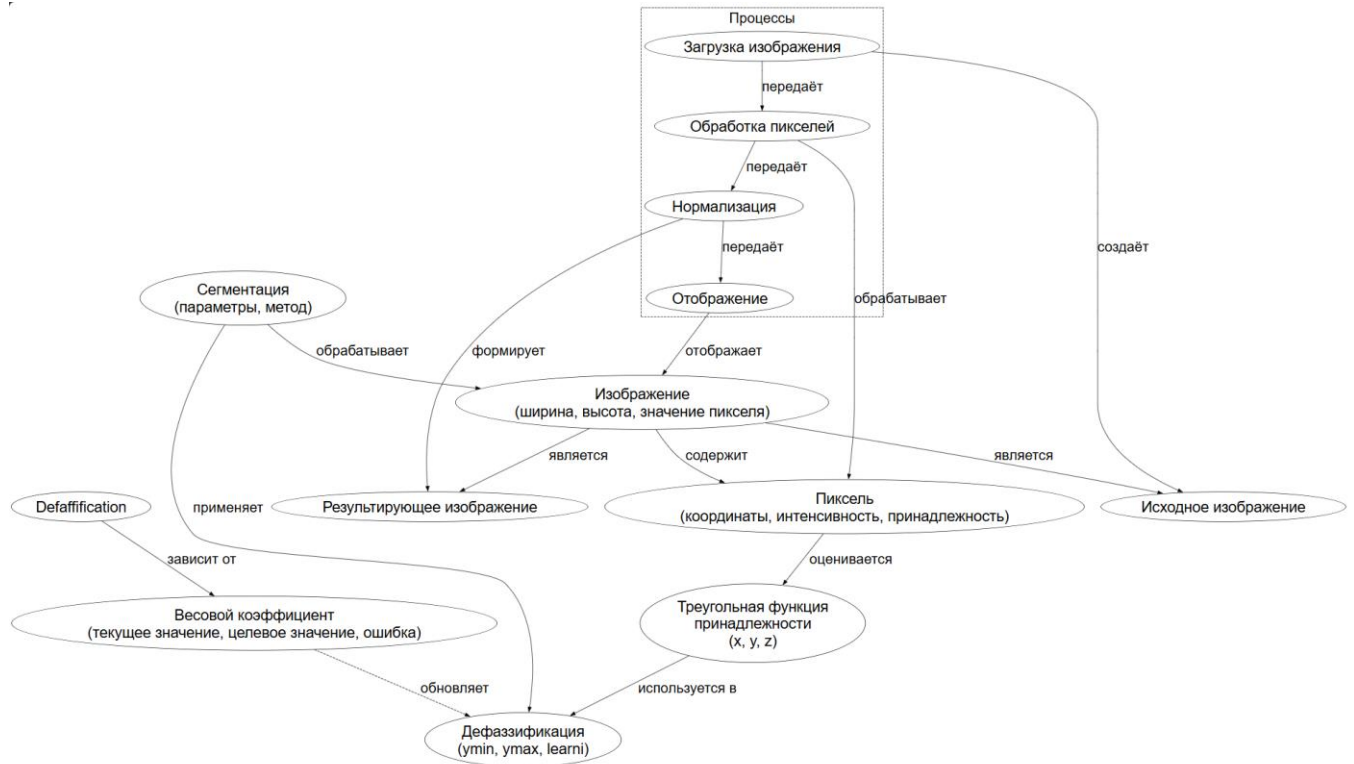


Рис. 2. Онтологическая модель ИАС сегментации изображений с функцией обучения

Онтологическая модель нейро-нечеткой сегментации изображений с функцией обучения состоит из набора понятий и отношений между ними. Ниже приведены таблицы описания онтологии объектов (табл. 1), онтологии атрибутов (табл. 2) и онтологии процессов (табл. 3).

Таблица 1. Онтология объектов

№	Название	Описание	Взаимосвязь
1.	Изображение (Image)	Двумерный массив пикселей, загружаемый из файла Подтипы: - Исходное изображение (OriginalImage): Входное изображение в градациях серого - Результирующее изображение (ResultImage): Выходное изображение после обработки	Пиксель (Pixel), Сегментация (Segmentation)
2.	Пиксель (Pixel)	Элемент изображения с определенной интенсивностью	Треугольная функция принадлежности
3.	Треугольная функция принадлежности (TriangularMembership Function)	Математическая функция для вычисления степени принадлежности пикселя к нечеткому множеству	Дефаззификация
4.	Дефаззификация (Defuzzification)	Процесс преобразования нечетких значений в четкое значение	Весовой коэффициент
5.	Весовой коэффициент (Learn _i)	Адаптивный параметр, обновляемый для минимизации ошибки сегментации	Ошибка
6.	Сегментация (Segmentation)	Процесс разделения изображения на фон и объект	Дефаззификация, Изображение

Таблица 2. Онтология атрибутов

№	Название объекта	Атрибут	Описание
1.	Изображение (Image)	Ширина (Width)	Количество пикселей по горизонтали
		Высота (Height)	Количество пикселей по вертикали
		Значение пикселя (PixelValue)	Интенсивность пикселя в градациях серого
2.	Пиксель (Pixel)	Координаты (Coordinates)	Позиция (i, j) в изображении
		Интенсивность (Intensity)	Значение от 0 до 255 (pixel_value)
		Принадлежность (Membership)	Степень принадлежности к нечеткому множеству (рассчитывается функцией triangular_membership)
3.	Треугольная функция принадлежности	x	Нижняя граница треугольника
		y	Вершина треугольника (максимальная принадлежность)
		z	Верхняя граница треугольника
4.	Дефаззификация (Defuzzification)	ymin	Минимальное выходное значение (0)
		ymax	Максимальное выходное значение (255)
		learn _i	Весовой коэффициент обучения
5.	Весовой коэффициент (Learn _i)	Текущее значение (CurrentValue)	Значение learn _i в диапазоне [0.001, 10]
		Целевое значение (TargetValue)	Ожидаемое значение для объекта (ytarget)

		Ошибка (Error)	Разница между $mar1$ и y_{target}
6.	Сегментация (Segmentation)	Параметры (Parameters)	Список параметров треугольных функций принадлежности ($triangular_params$)
		Метод (Method):	Нечеткая логика с дефаззификацией

Таблица 3. Онтология процессов

№	Название	Описание	Взаимосвязь
1.	Загрузка изображения	Вход: Файл изображения Выход: $OriginalImage$	Исходное изображение ($OriginalImage$)
2.	Обработка пикселей	Вход: $OriginalImage$ Выход: $ResultImage$ Действия: • Вычисление принадлежности ($triangular_membership$) • Дефаззификация ($defuzzification$) • Обновление $learn_i$ ($update_learn_i$)	Пиксель (Pixel)
3.	Нормализация	Вход: $ResultImage$ (тип $float32$) Выход: $ResultImage$ (тип $uint8$, диапазон 0–255)	Результирующее изображение ($ResultImage$)
4.	Отображение	Вход: $OriginalImage$, $ResultImage$. Действие: Вывод на экран с помощью $cv2.imshow$	Изображение ($Image$)

Для наглядности приведем псевдокод для моделирования онтологической модели сегментации изображений:

1. // Функция вычисления треугольной принадлежности
2. функция `TriangularMembership(s, x, y, z)`:
3. если $s < x$ ИЛИ $s > z$:
4. вернуть 0
5. иначе если $x \leq s \leq y$:
6. вернуть $(s - x) / (y - x)$
7. иначе:
8. вернуть $(z - s) / (z - y)$
- 9.
10. // Функция дефаззификации
11. функция `Defuzzification(memberships, ymin=0, ymax=255, learni=1)`:
12. $n = \text{длина}(\text{memberships})$
13. $\text{term} = \text{сумма}(\text{memberships}) / (n * \text{learni})$
14. $\text{result} = \text{term} * (\text{ymax} - \text{ymin}) + \text{ymin}$
15. вернуть `ограничить(result, 0, 255)`
- 16.
17. // Функция обновления весового коэффициента
18. функция `UpdateLearni(learni, mar1, ytarget, w=0.004, t=0.01)`:
19. если `абсолютная_разница(mar1, ytarget) > t`:
20. $\text{learni_new} = \text{learni} + w * (\text{ytarget} - \text{mar1})$
21. вернуть `ограничить(learni_new, 0.001, 10)`
22. вернуть `learni`
- 23.
24. // Основной алгоритм
25. начало:
26. загрузить изображение в градациях серого как `image`
- 27.
28. задать параметры треугольных функций:
29. $\text{params} = [(0, 85, 170), (85, 170, 255)]$
- 30.
31. создать пустое изображение `result_image` такого же размера, как `image`
- 32.
33. установить начальные параметры:
34. $\text{learni} = 1$
35. $\text{ytarget} = 170$
- 36.

```
37. для каждого пикселя (i,j) в изображении:
38.     pixel_value = image[i,j]
39.
40.     memberships = []
41.     для каждого набора параметров в params:
42.         добавить TriangularMembership(pixel_value, params) в memberships
43.
44.     mar1 = Defuzzification(memberships, 0, 255, learni)
45.     learni = UpdateLearn(i, mar1, ytarget)
46.     result_image[i,j] = mar1
47.
48. нормализовать result_image в диапазон [0, 255]
49. преобразовать result_image в 8-битный формат
50.
51. показать оригинальное изображение
52. показать сегментированное изображение
53. конец
```

Разработанная онтологическая модель нейро-нечеткой сегментации изображений с функцией обучения представляет собой структурированное и формализованное описание системы, включающее ключевые объекты, их атрибуты и процессы взаимодействия. Предложенные онтологии объектов, атрибутов и процессов обеспечивают четкое представление взаимосвязей между элементами системы, такими как изображение, пиксели, треугольные функции принадлежности, дефаззификация и адаптивный весовой коэффициент. Это позволяет не только концептуально описать алгоритм сегментации, но и заложить основу для его дальнейшей программной реализации и анализа. Псевдокод, приведенный в разделе, отражает практическую реализацию онтологической модели, демонстрируя последовательность шагов обработки данных и их соответствие теоретическим положениям. Таким образом, онтологическая модель служит связующим звеном между математической основой метода и его прикладным применением, подчеркивая системный подход к решению задачи сегментации изображений и потенциал для масштабирования алгоритма в более сложных задачах компьютерного зрения.

ЭКСПЕРИМЕНТЫ

Были проведены две серии экспериментов (с разными изображениями) для наглядного представления результатов работы предложенной модели. На рис. 2–4 представлена первая серия экспериментов, на рис. 5–7 – вторая серия экспериментов. Для этих серий экспериментов была поставлена цель – сегментировать объект от остального фона. В ходе экспериментов также было проведено сравнение разработанной модели с модифицированным методом отношения площадей; модели со стандартным методом отношения площадей.



Рис. 2. а) Оригинальное изображение; б) в градации серого



Рис. 3. Изображение, обработанное с помощью:
а) метода отношения площадей;
б) модифицированного метода отношения площадей

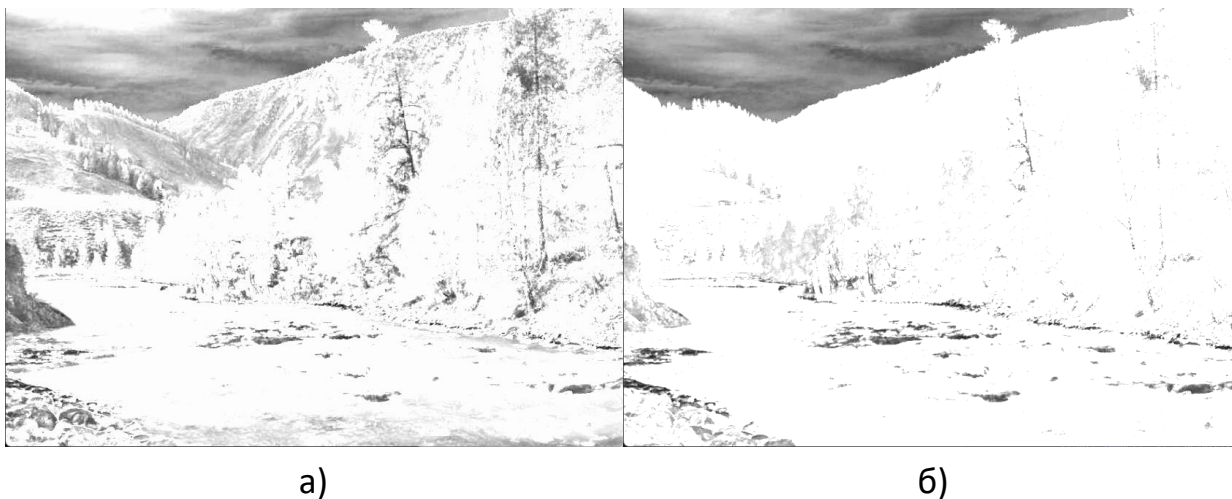


Рис. 4. Изображение, обработанное с помощью нейро-нечеткой системы обучения на основе: а) метода отношения площадей; б) модифицированного метода отношения площадей

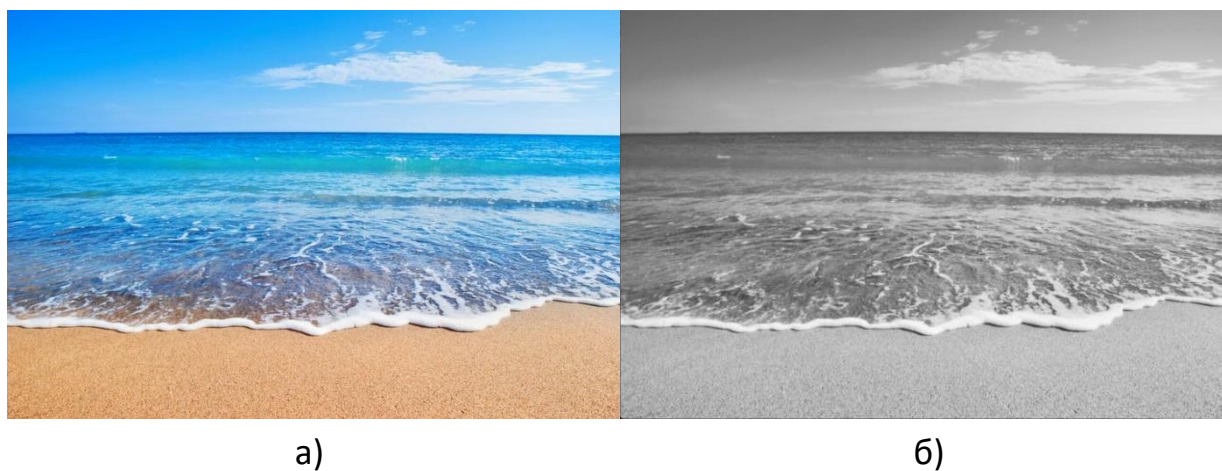


Рис. 5. а) Оригинальное изображение; б) в градации серого



а)

б)

Рис. 6. Изображение, обработанное с помощью:
а) метода отношения площадей;
б) модифицированного метода отношения площадей



Рис. 7. Изображение, обработанное с помощью нейро-нечеткой системы обучения на основе: а) метода отношения площадей;
б) модифицированного метода отношения площадей

На основе полученных результатов был рассчитан коэффициент Дайса по формуле

$$\text{Dice} = \frac{2|A \cap B|}{|A| + |B|},$$

где $|A \cap B|$ – количество общих элементов в наборах A и B , $|A|$ – количество элементов в наборе A , $|B|$ – количество элементов в наборе B . Он позволяет количественно оценить степень пересечения элементов в исследуемых выборках. Этот показатель, варьирующийся от 0 до 1, где 1 означает полное совпадение, а 0 – полное отсутствие общих элементов, оказался полезным инструментом для ин-

терпретации данных. Эксперименты включали обработку реальных наборов данных (интенсивность пикселей), что обеспечило надежность и применимость полученных выводов в дальнейшем анализе. Результаты расчетов представлены в табл. 4.

Таблица 4. Результаты экспериментов

		MAR	MAR2	MAR + обучение	MAR2 + обучение
Dice	1-й эксперимент	0.4255	0.7776	0.7558	0.9226
	2-й эксперимент	0.4375	0.5468	0.4545	0.6635

На основании результатов экспериментов можно сделать следующие выводы.

1. Для первого эксперимента значение коэффициента Дайса при использовании метода MAR составило 0.4255, а при методе MAR2 — 0.7776. При добавлении обучения (MAR+обучение) значение увеличилось до 0.7558, а с комбинацией MAR2+обучение достигло максимального значения 0.9226. Это подтверждает, что обучение значительно улучшает результаты, особенно в сочетании с методом MAR2.

2. Для второго эксперимента начальное значение коэффициента Дайса при методе MAR равно 0.4375, а при MAR2 – 0.5468. С добавлением обучения (MAR+обучение) результат повысился до 0.4545, а при использовании MAR2+обучение достиг наивысшего значения 0.6635 среди всех вариантов для этого эксперимента.

Общий анализ показал, что метод MAR2 в комбинации с обучением обеспечивает наилучшие значения коэффициента Дайса по сравнению с традиционным MAR. Наибольший прирост эффективности наблюдался в первом эксперименте. Во втором эксперименте прирост также значительный, но менее выраженный.

Таким образом, что комбинация метода MAR2 с обучением является наиболее эффективной стратегией для повышения значений коэффициента Дайса в обоих проведенных экспериментах.

ЗАКЛЮЧЕНИЕ

Разработанный нейро-нечеткий алгоритм для высокоскоростной сегментации изображений в градациях серого с функцией обучения демонстрирует значительные преимущества по сравнению с традиционными методами. Экспериментальные результаты, полученные на изображениях формата 1024x720, подтвердили, что использование модифицированного метода отношения площадей с адаптивным обучением весового коэффициента позволяет достичь высокой точности сегментации и улучшенной контрастности объектов при минимальных вычислительных затратах. Сравнение с базовым методом отношения площадей установило превосходство предложенного подхода в задачах выделения объектов, таких как сегментация воды на фоне. Эти результаты подтверждают перспективность применения искусственного интеллекта и нечеткой логики в области компьютерного зрения. В дальнейшем планируется расширение исследования путем интеграции алгоритма с многослойными нейронными сетями и тестирования на более сложных наборах данных, что может повысить его универсальность и эффективность в реальных приложениях.

Благодарности

Работа выполнена при поддержке Министерства науки и высшего образования РФ в рамках работ по Государственному заданию № 075-03-2025-526.

СПИСОК ЛИТЕРАТУРЫ

1. *Bobyry M.V., Milostnaya N.A., Bulatnikov V.A.* The fuzzy filter based on the method of areas' ratio // *Applied Soft Computing*, V. 117, 2022, 108449, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2022.108449>
2. *Бобырь М.В., Архипов А.Е., Милостная Н.А.* МЕТОД РАСЧЕТА КАРТЫ ГЛУБИН НА ОСНОВЕ МЯГКИХ ОПЕРАТОРОВ // *Системы и средства информатики*. 2019. Т. 29. № 2. С. 71–84. <https://doi.org/10.14357/08696527190207>
3. *Lee D.H., Chen P.Y., Yang F.J. et al.* "High-Efficient Low-Cost VLSI Implementation for Canny Edge Detection" // *Journal of Information Science & Engineering*, 2020. Vol. 36, no. 3. P. 34–57.

4. Koohi H., Kiani K. User Based Collaborative Filtering Using Fuzzy C-Means // Measurement. 2016. V. 91. P. 134–139. <https://doi.org/10.1016/j.measurement.2016.05.058>
5. Yang Q., Sun L. A Fuzzy Complementary Kalman Filter Based on Visual and IMU Data for UAV Landing // Intern. Journal for Light and Electron Optics. 2018. V. 173. P. 279–291. <https://doi.org/10.1016/j.ijleo.2018.08.011>.
6. S. Eti, S. Yüksel, H. Dinçer. A machine learning and fuzzy logic model for optimizing digital transformation in renewable energy: Insights into industrial information integration // Journal of Industrial Information Integration. 2024. V. 42. P. 100734. <https://doi.org/10.1016/j.jii.2024.100734>. EDN MFEAPI.
7. Romanov A.A., Filippov A.A., Yarushkina N.G. Adaptive Fuzzy Predictive Approach in Control // Mathematics. 2023. V. 11, No. 4. P. 875. <https://doi.org/10.3390/math11040875>. EDN LRGVQT.
8. Bobyr M., Bondarenko B., Malyshev A. High-speed Fuzzy Inference Machine Learning Device Based on Single-Layer Area Ratio Defuzzifier // Intelligence Enabled Research: Proceedings of the 2024 Sixth Doctoral Symposium on Intelligence Enabled Research (DoSIER 2024), Dhupguri, Jalpaiguri, West Bengal, India, November 28–29, 2024. Jalpaiguri, West Bengal, India, 2025. P. 15–25. EDN FXFCKM.
9. Зимичев Е.А., Казанский Н.Л., Серафимов П.Г. Пространственная классификация путем интеграции изображений с использованием метода кластеризации k-means ++ // Компьютерная оптика. 2014. Т. 38, № 2. С. 281–286. <https://doi.org/10.18287/01342452-2014-38-2-281-286>
10. Pereyra M., McLaughlin S. Fast Unsupervised Bayesian Image Segmentation with Adaptive Spatial Regularisation // IEEE Transactions on Image Processing, 2017. V. 26, No. 6. P. 2577–2587. <https://doi.org/10.1109/TIP.2017.2675165>

NEURO-FUZZY IMAGE SEGMENTATION WITH LEARNING FUNCTION

M. V. Bobyr¹ [0000-0002-5400-6817], B. A. Bondarenko² [0000-0001-5415-9015]

^{1, 2}South-West State University, Kursk, 305040, Russia

¹maxboby@gmail.com, ²sikersinko@gmail.com

Abstract

This paper presents a neuro-fuzzy algorithm for high-speed grayscale image segmentation based on a modified defuzzification method using triangular membership functions. The aim of the study is to analyze the effect of simplifying the defuzzification formula on the accuracy and contrast of object selection. The proposed approach includes adaptive learning of the weight coefficient, which allows dynamically adjusting the defuzzification process depending on the target values. The paper compares the basic method of averaging membership values and a modified version taking into account nonlinear weights. Experiments conducted on 1024x720 images demonstrate that the developed algorithm provides high segmentation accuracy and improved object contrast with minimal computational costs. The results confirm the superiority of the proposed method over traditional approaches, emphasizing the prospects for applying artificial intelligence in computer vision problems.

Keywords: IAS, neuro-fuzzy algorithm, image segmentation, defuzzification, artificial intelligence, area ratio method.

REFERENCES

1. Bobyr M.V., Milostnaya N.A., Bulatnikov V.A. The fuzzy filter based on the method of areas' ratio // Applied Soft Computing, Volume 117, 2022, 108449, ISSN 1568-4946, <https://doi.org/10.1016/j.asoc.2022.108449>
2. Bobyr' M.V., Arkhipov A.Ye., Milostnaya N.A. Metod rascheta karty glubin na osnove myagkikh operatorov // Sistemy i sredstva informatiki. 2019. T. 29. № 2. S. 71–84, <https://doi.org/10.14357/08696527190207>
3. Lee D.H., Chen P.Y., Yang F.J. et al. "High-Efficient Low-Cost VLSI Implementation for Canny Edge Detection" // Journal of Information Science & Engineering, 2020. Vol. 36, No. 3. P. 34–57.

4. Koohi H., Kiani K. User Based Collaborative Filtering Using Fuzzy C-Means // Measurement. 2016. V. 91. P. 134–139.
<https://doi.org/10.1016/j.measurement.2016.05.058>
 5. Yang Q., Sun L. A Fuzzy Complementary Kalman Filter Based on Visual and IMU Data for UAV Landing // Intern. Journal for Light and Electron Optics. 2018. V. 173. P. 279–291. <https://doi.org/10.1016/j.ijleo.2018.08.011>.
 6. S. Eti, S. Yüksel, H. Dinçer. A machine learning and fuzzy logic model for optimizing digital transformation in renewable energy: Insights into industrial information integration // Journal of Industrial Information Integration. 2024. Vol. 42. P. 100734. <https://doi.org/10.1016/j.jii.2024.100734>. EDN MFEAPI.
 7. Romanov A.A., Filippov A.A., Yarushkina N.G. Adaptive Fuzzy Predictive Approach in Control // Mathematics. 2023. Vol. 11, No. 4. P. 875.
<https://doi.org/10.3390/math11040875>. EDN LRGVQT.
 8. Bobyr M., Bondarenko B., Malyshev A. High-speed Fuzzy Inference Machine Learning Device Based on Single-Layer Area Ratio Defuzzifier // Intelligence Enabled Research: Proceedings of the 2024 Sixth Doctoral Symposium on Intelligence Enabled Research (DoSIER 2024), Dhupguri, Jalpaiguri, West Bengal, India, November 28–29, 2024. Jalpaiguri, West Bengal, India, 2025. P. 15–25. EDN FXFCKM.
 9. Zimichev E.A., Kazanskiy N.L., Serafimov P.G. Prostranstvennaya klassifikatsiya putem integratsii izobrazheniy s ispol'zovaniyem metoda klasterizatsii k-means ++ // Komp'yuternaya optika. 2014. T. 38, No. 2. S. 281–286.
<https://doi.org/10.18287/01342452-2014-38-2-281-286>
 10. Pereyra M., McLaughlin S. Fast Unsupervised Bayesian Image Segmentation with Adaptive Spatial Regularisation // IEEE Transactions on Image Processing, 2017. Vol. 26, No. 6. P. 2577–2587. <https://doi.org/10.1109/TIP.2017.2675165>
-

СВЕДЕНИЯ ОБ АВТОРАХ



БОБЫРЬ Максим Владимирович – 1978 года рождения, учился в Курском государственном техническом университете (ныне – Юго-Западный государственный университет). В 2012 году защитил диссертацию на соискание доктора технических наук по специальности 05.13.06 Автоматизация и управление технологическими процессами и производствами. В настоящее время работает профессором на кафедре программной инженерии. Является председателем диссертационного совета по специальности 5.12.4 Когнитивное моделирование. Область научных интересов: адаптивные нейро-нечеткие системы вывода, цифровая обработка изображений и распознавание образов, машинное обучение, стереозрение.

Maksim Vladimirovich BOBYR – born in 1978, studied at Kursk State Technical University (now South-West State University). In 2012, he defended his dissertation for the degree of Doctor of Technical Sciences in specialty 05.13.06 Automation and Control of Technological Processes and Production. Currently, he works as a professor at the Department of Software Engineering. He is the chairman of the dissertation council for specialty 5.12.4 Cognitive Modeling. Research interests: adaptive neuro-fuzzy inference systems, digital image processing and pattern recognition, machine learning, stereo vision.

email: maxbobyrgmail.com

ORCID: 0000-0002-5400-6817



БОНДАРЕНКО Богдан Андреевич – 1997 года рождения, учится на 4 курсе аспирантуры Юго-Западного государственного университета на кафедре программной инженерии по специальности 05.13.06 Автоматизация и управление технологическими процессами и производствами. Область научных интересов: обучающие нечетко-логические системы, распознавание образов.

Bogdan Andreevich BONDARENKO – born in 1997, a 4th-year postgraduate student at the Southwestern State University, Department of Software Engineering, specialty 05.13.06 Automation and Control of Technological Processes and Production. Research interests: training fuzzy logic systems, pattern recognition.

email: sikersinko@gmail.com

ORCID: 0000-0001-5415-9015

Материал поступил в редакцию 26 марта 2025 года

УДК 004.4

ПРОГРАММНЫЙ МОДУЛЬ ФОРМИРОВАНИЯ ЦИФРОВОГО МАТЕМАТИЧЕСКОГО ПРОСТРАНСТВА НА ОСНОВЕ ГРАФОВ ЗНАНИЙ

В. И. Гурьянов¹ [0009-0008-9126-7216], А. М. Елизаров² [0000-0003-2546-6897]

^{1, 2}Казанский федеральный университет, г. Казань, 420008, Россия

²Национальный исследовательский центр (НИЦ) «Курчатовский институт»,
Москва, 123182, Россия

¹vadim.guryanov.2004@mail.ru, ²amelizarov@gmail.com

Аннотация

Современное информационное пространство содержит множество данных, однако они зачастую слабо структурированы, трудно находимы и не всегда корректны. Это создаёт дополнительные трудности при исследованиях, поэтому в настоящее время формируются цифровые пространства научных знаний, в частности, на основе графов знаний.

Для обеспечения качества информации такие графы часто наполняются данными вручную, что требует больших затрат времени. Поэтому создание инструмента, предоставляющего возможность автоматического наполнения графа данными, а также обеспечивающего контроль их качества, позволит упростить и ускорить процесс формирования цифровых пространств научных знаний.

Предложены методы автоматизации наполнения графа данными, обеспечивающие параллельный контроль их целостности. На основе предложенных методов разработан программный модуль, описаны механизмы его функционирования и его архитектура.

Ключевые слова: цифровое пространство научных знаний, формирование цифровых пространств научных знаний, графы знаний, автоматизация построения графов знаний.

ВВЕДЕНИЕ

Современное информационное пространство содержит множество данных из разнообразных предметных областей, однако эта информация часто слабо

структурирована, трудно находима и к тому же может оказаться некорректной. Всё это значительно осложняет поиск и обработку необходимой информации. Поэтому актуальной задачей является создание цифровых пространств научных знаний, позволяющих систематизировать доступ к данным, упростить их обработку и использование, а также гарантировать их корректность (см. также [1, 2]).

Проблема обеспечения исследователей качественными, проверенными и непротиворечивыми данными давно изучается многими исследователями как в России, так и в других странах. В зарубежных публикациях этот вопрос обычно рассматривается в рамках более общего понятия – разработки исследовательских инфраструктур [3–5]. Отечественные авторы часто рассматривают этот вопрос с позиции создания цифровых пространств научных знаний [1, 2, 6, 7]. Оба этих понятия подразумевают формирование некоторой компьютерной среды, объединяющей научные знания, а также предоставляющей дополнительные инструменты для их поиска и использования. В последнее время для создания такой среды изучают возможности применения онтологий и графов знаний (например, [8–10]).

Как разновидность научных знаний, в науке активно изучают математическое знание. Построение цифровых математических пространств и графов математических знаний рассматривают многие авторы, например, [8, 11–14]. Это актуально, поскольку ещё не создан граф, полностью охватывающий всё математическое знание.

Существует множество подходов к созданию графов знаний, каждый из которых сопряжён со своими преимуществами и недостатками. Первый подход – это автоматизированное формирование графа на основе программного извлечения и преобразования данных. Примером использования такого метода является граф DBpedia [15], построенный путём преобразования информации, собранной в рамках проекта Wikipedia, в структурированные данные. Преимуществом программного наполнения этого графа является возможность обработки большого объёма данных. Однако очевидны и недостатки: для применения метода необходимы веб-ресурсы или база данных, уже содержащих необходимую информа-

цию. Кроме того, при сборе информации из разных источников для каждого может потребоваться разработка отдельного сборщика данных, и существенным становится вопрос доверия к извлекаемым данным.

Другим подходом является предоставление доступа к открытому редактированию графа знаний широкой группе пользователей. Например, таким способом был реализован проект Wikidata [16]. Преимуществом такого способа является постепенное наполнение графа данными без необходимости дополнительного вмешательства со стороны команды проекта: постепенно пользователи будут сами интегрировать в граф информацию из различных источников. Недостатком является необходимость дополнительного привлечения сотрудников для наполнения графа, стоит также вопрос доверия к вносимым данным и даже имеется вероятность злого умысла в виде внесения заведомо ложной информации.

Третий подход – это ручное наполнение графа данными. В этом случае некоторая группа квалифицированных специалистов наполняет граф знаний проверенными данными. Например, таким способом были реализованы проекты MathAlgoDB [17] и OntoMathPro [12, 13]. Названный подход схож со вторым, однако круг редакторов графа заведомо ограничивается доверенными экспертами. Преимуществами являются высокое качество и структурированность данных. Однако такой подход приводит к большим временным трудозатратам. В процессе ручного создания графа также появляется вероятность возникновения ошибок из-за человеческого фактора, что может негативно отразиться на итоговом качестве.

Так как качество данных является приоритетом при создании графов научных знаний, в процессе их формирования часто применяют методы ручного наполнения данными [17].

Таким образом, можно утверждать, что существует потребность в инструменте, предоставляющем возможность не только интеграции систем автоматизации сбора информации и её добавления в граф, но и параллельного контроля качества вносимых данных. По этой причине нами был разработан инструмент, обладающий следующими свойствами:

- возможность расширения функционала;
- контроль типа вносимых объектов;
- возможность автоматического назначения типа объекта;
- применимость к уже построенному графу;

- наличие функций для вызова SPARQL-запросов из программного кода;
- построение подграфа на основе SPARQL-запроса;
- открытый исходный код.

ИНСТРУМЕНТЫ, ИСПОЛЬЗУЕМЫЕ ПРИ ПОСТРОЕНИИ ГРАФОВ ЗНАНИЙ

В настоящее время существует несколько инструментов, реализующих различные реализации поставленной задачи, например Protégé [18]; RDFLib [19]; SPARQLWrapper [20]. Некоторое сравнение этих инструментов представлено в табл. 1.

Таблица 1. Сравнение альтернативных решений

	Protégé	RDFLib	SPARQLWrapper
Возможность взаимодействия через язык программирования	–	+	+
Взаимодействие с базой данных	–	–	+
Контроль целостности данных	+	–	–
Открытый исходный код	+	+	+

Каждый из представленных инструментов решает различные задачи, но не соответствует всем необходимым критериям (см. табл. 1).

Protégé – редактор онтологий с открытым исходным кодом. Он предоставляет множество инструментов для редактирования онтологий и контроля целостности данных. При этом вся работа осуществляется через графический интерфейс, что, хотя и является удобным для ручной модификации графа, но накладывает значительные ограничения на организацию процесса автоматизации работы.

RDFLib – библиотека для языка Python с открытым исходным кодом. Содержит функции для работы с RDF-документами и поддерживает многие часто используемые стандарты, такие как RDF, RDFS, OWL, FOAF. Однако функционал этой библиотеки подразумевает полную загрузку графа в оперативную память, что неприемлемо при работе с крупными графами.

SPARQLWrapper – библиотека для языка Python, разработанная для взаимодействия с базой данных через точку доступа SPARQL. Это позволяет работать напрямую с графом и не требует загрузки всех данных в оперативную память. Однако эта библиотека дает только функционал для выполнения текстового SPARQL-запроса и не содержит какого-либо другого функционала взаимодействия с графом, например проверки целостности.

ЛОГИКА КОНТРОЛЯ ЦЕЛОСТНОСТИ ДАННЫХ

Основными элементами взаимодействия между модулями созданной нами программы являются Triple и TripleItem.

Triple представляет собой реализацию RDF-триплета и состоит из субъекта, предиката и объекта, каждый из которых является элементом типа TripleItem.

В соответствии со стандартом RDF [21] «субъект» и «предикат» должны представлять собой некий идентификатор (IRI или идентификатор пустого узла), а «объект» может быть как идентификатором, так и просто значением.

Для контроля целостности данных все объекты были разделены на два типа – Entity (сущность) и Predicate (предикат). Каждому из них соответствует отдельная таблица в реляционной базе данных, содержащая Id и IRI объекта. Объект может быть добавлен в граф только после того, как будет внесён в соответствующую ему таблицу (регистрация).

Решение использовать реляционную базу данных для определения типа объектов было принято по той причине, что применение для решения рассматриваемой задачи только графовой базы данных потребует либо рекурсивного обхода части узлов, либо создания отдельного графа, содержащего только структурную информацию. Оба этих метода менее оптимальны, чем использование реляционной базы данных.

Для регистрации объекта существует специальная функция. Ей нужно передавать его тип – Entity или Predicate. Имеется также возможность дополнительно

передать Id или IRI, которые будут обозначать создаваемый объект. Сначала проводится проверка существования объекта с таким IRI/Id в соответствующей таблице базы данных. Если объект существует, то информация о нём извлекается из базы данных. Если IRI не был передан, то он формируется на основе переданного из базы данных или нового Id. Далее объект добавляется в граф.

В граф разрешено добавлять только связи вида:

- Сущность – Предикат – Значение (например, Теорема Ферма – Сформулирована – 1637 год);
- Сущность – Предикат – Сущность (например, Алгебра – Относится к – Математика);
- Предикат – Предикат – Значение (Относится к – Добавлен – Дата добавления);
- Предикат – Предикат – Предикат (Относится к – Класс – Predicate).

Исключением является возможность установки отношения эквивалентности между предикатом и сущностью. Это может быть полезным, если необходимо указать связь между предикатом и реальными объектом или понятием. Например, для предиката «Имя», устанавливающего связь между человеком и его именем, может потребоваться организация связи с понятием «Имя», объясняющим на основе связей, что такое имя имеется у человека.

Добавление триплета в граф реализуется специальной функцией. По умолчанию для переданного RDF-триплета проверяется, соответствует ли он одному из четырёх ранее описанных видов связей, а для каждого его элемента определяется, был ли он ранее зарегистрирован.

Блок-схема алгоритма проверки целостности показана на рис. 1. Если требования к качеству данных слишком строгие, то функция сама может проводить регистрацию недостающих объектов. Для этого достаточно передать этой функции соответствующий параметр. Вывод о том, какой тип должен быть присвоен незарегистрированному элементу RDF-триплета, основан на разрешённых связях.

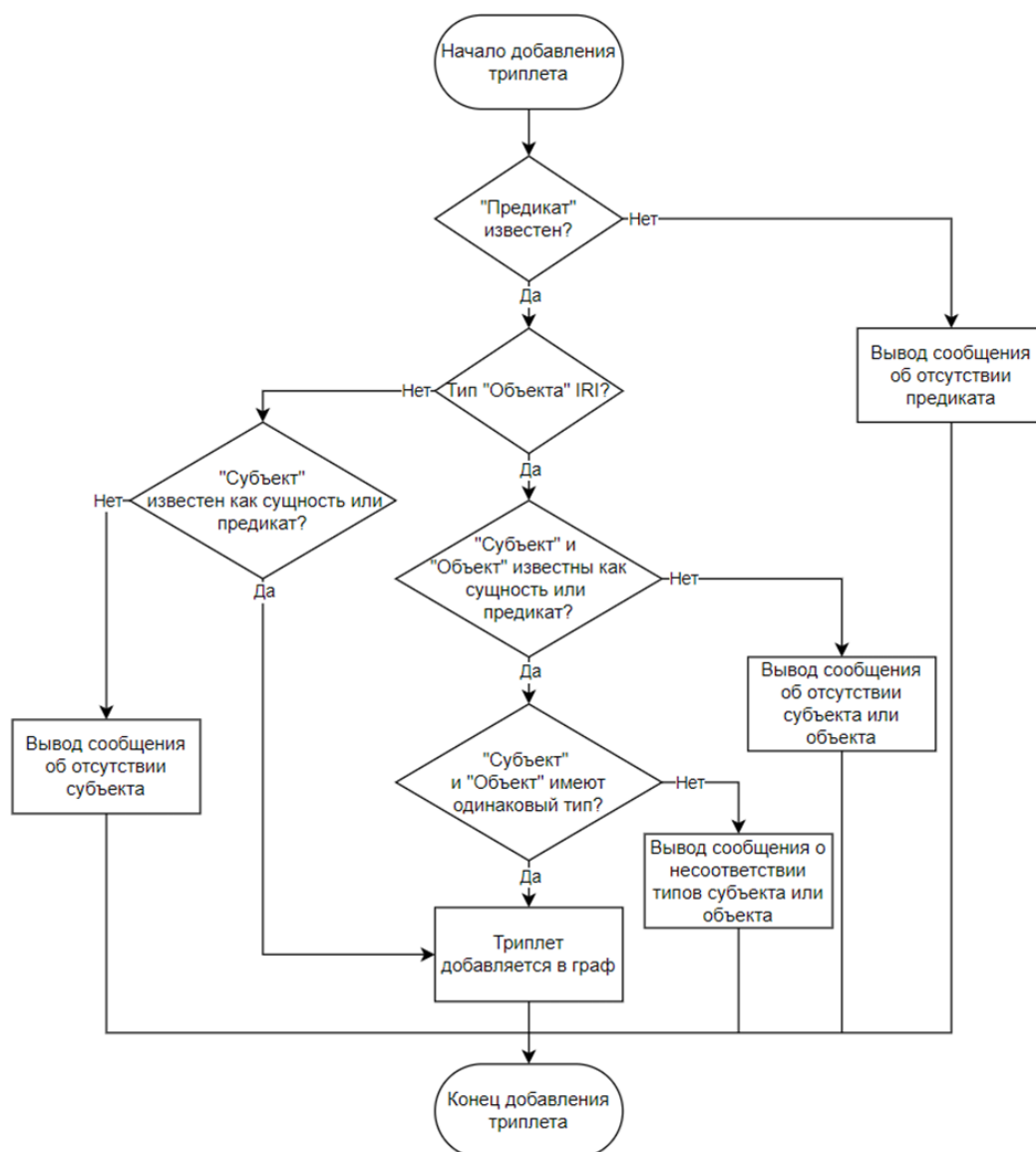


Рис. 1. Алгоритм добавления триплета с проверкой целостности

Рассмотрим правила автоматической регистрации недостающих объектов. Пусть RDF-триплет представляет собой тройку (s, p, o) . Обозначим множество зарегистрированных объектов через R , множества сущностей, предикатов и значений – соответственно через E , P и L . Тогда:

- «Предикат» триплета не зарегистрирован, следовательно, его тип – Predicate ($p \notin R \Rightarrow p \in P$).
- «Субъект» («Объект») не зарегистрирован, но «Объект» («Субъект») зарегистрирован, следовательно, тип «Субъекта» («Объекта») равен типу «Объекта» («Субъекта»)

$$(s \notin R, o \in E \Rightarrow s \in E; s \notin R, o \in P \Rightarrow s \in P; o \notin R, s \in E$$

$$\Rightarrow o \in E; o \notin R, s \in P \Rightarrow o \in P).$$

- «Субъект» и «Объект» не зарегистрированы, следовательно, их тип – Entity ($s \notin R, o \notin E \Rightarrow s \in E, o \in E$).
- «Субъект» не зарегистрирован, а «Объект» является значением, следовательно, «Субъект» – это Entity ($s \notin R, o \in L \Rightarrow s \in E$).

Последние два условия выполнены на основе того, что операция добавления сущности в граф статистически гораздо более частая, чем добавление предиката. Блок-схема алгоритма для этого случая показана на рис. 2.

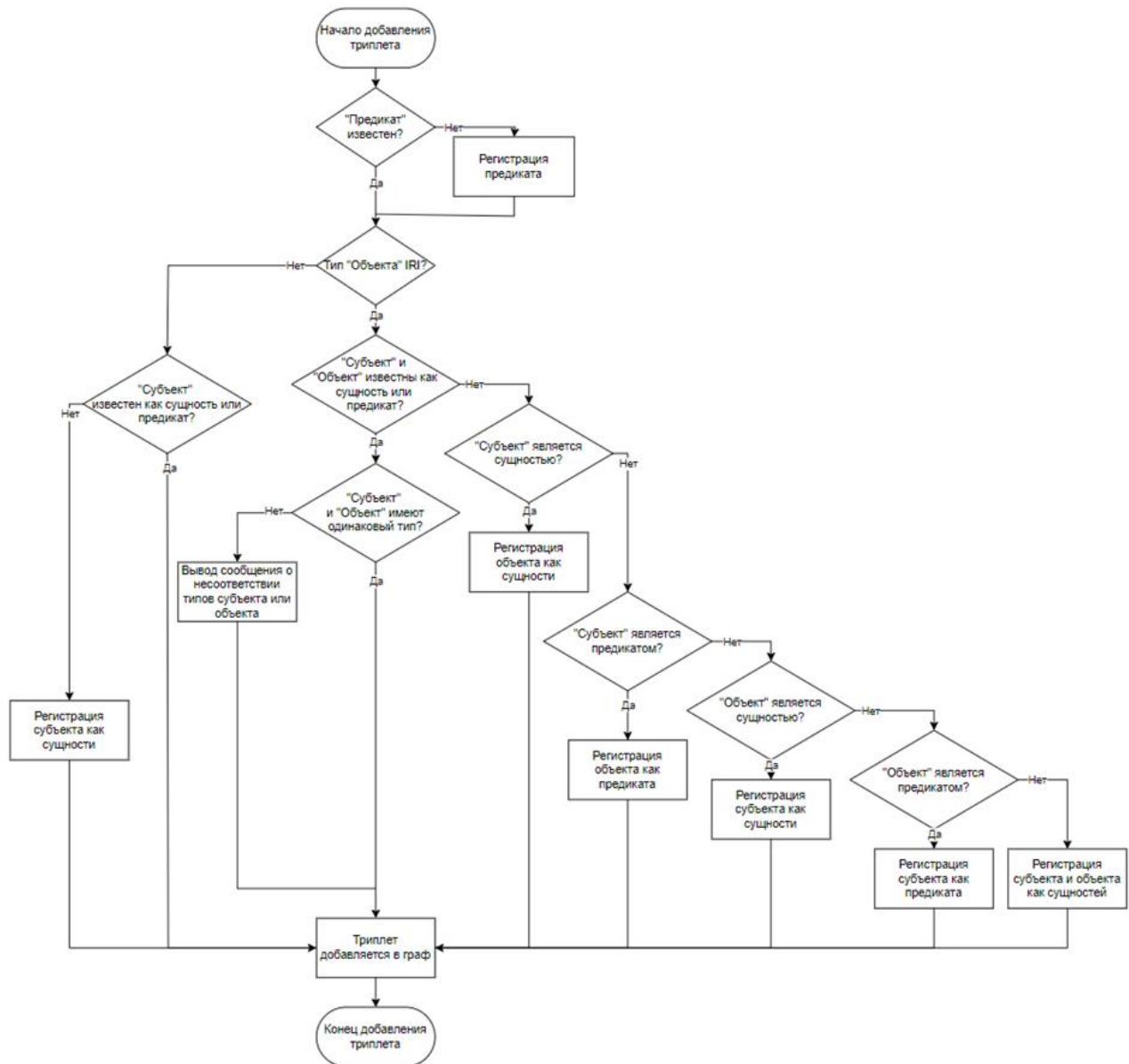


Рис. 2. Алгоритм добавления триплета с автоматической регистрацией

АРХИТЕКТУРА РАЗРАБОТАННОГО МОДУЛЯ

Для взаимодействия с системой управления данными Virtuoso Universal Server [22] был создан класс, содержащий функции-конструкторы SPARQL-запросов. Конструкторы были разделены на два типа в зависимости от выполняемых операций: вносящие изменения в базу данных (операции записи) и применяемые для получения информации из базы данных (операция чтения). Сконструированные запросы выполняются с применением библиотеки SPARQLWrapper. На данный момент поддерживаются следующие операции:

- SELECT – для получения данных из графа;
- CREATE GRAPH – для создания графа;
- DROP GRAPH – для удаления графа;
- INSERT – для добавления триплетов в граф;
- DELETE – для удаления триплетов из графа.

Сконструированные запросы на изменение графа выполняются не сразу, а только после вызова специальной функции Commit, что позволяет объединить несколько запросов в один и уменьшить время ожидания их выполнения.

В зависимости от цели результат SELECT-запроса можно получить в формате JSON или в виде массива RDF-триплетов. Библиотека SPARQLWrapper, выполняющая сформированный SPARQL-запрос, возвращает значения в формате JSON. Если результат требуется получить в виде массива RDF-триплетов, то он формируется на основе WHERE и OPTIONAL части запросов. Для этого каждая переменная в WHERE и OPTIONAL части запросов заменяется её значением, возвращённым в формате JSON.

Получение результата SPARQL SELECT-запроса в виде массива RDF-триплетов принципиально важно, так как позволяет получить подграф в формате, поддерживаемом другими разработанными модулями, такими как модуль визуализации и модуль взаимодействия с графом.

Взаимодействие с графами проводится через специальный класс который содержит функции, объединяющие взаимодействие с реляционной и графовой базами данных. Основными функциями этого класса являются

- регистрация объекта;
- добавление триплета в граф;
- копирование графа.

Функции регистрации объекта и добавления триплета в граф являются реализацией описанной выше логики контроля целостности данных.

Функция копирования графа позволяет с помощью конструктора SPARQL-запросов получить полный граф по переданной точке доступа SPARQL и построить на его основе новый граф вызовом предыдущих двух функций. Для этого сначала регистрируются все предикаты триплетов, а тип остальных объектов определяется на основе их связей по ранее описанному алгоритму. Эта функция позволяет интегрировать ранее созданные графы и применять к ним разработанную логику контроля целостности данных.

РЕЗУЛЬТАТЫ

Разработанный программный модуль был применён для решения ряда задач, описанных ниже. С сайта Казанского федерального университета была собрана информация о сотрудниках Института информационных технологий и интеллектуальных систем (ИТИС) и записана в граф. На основе полученного списка был проведён поиск публикаций сотрудников ИТИС в Научной электронной библиотеке Elibrary (<https://www.elibrary.ru/>). Полученная информация также была записана в граф и связана с ранее полученными данными обо всех сотрудниках института ИТИС. Часть построенного графа представлена на рис. 3.

Кроме того, разработанный программный модуль был успешно применён для работы с онтологией OntoMathPro (<https://github.com/CLLKazan/OntoMathPro/>) (см. также [12, 13]), что даёт больше возможностей взаимодействия с ней.

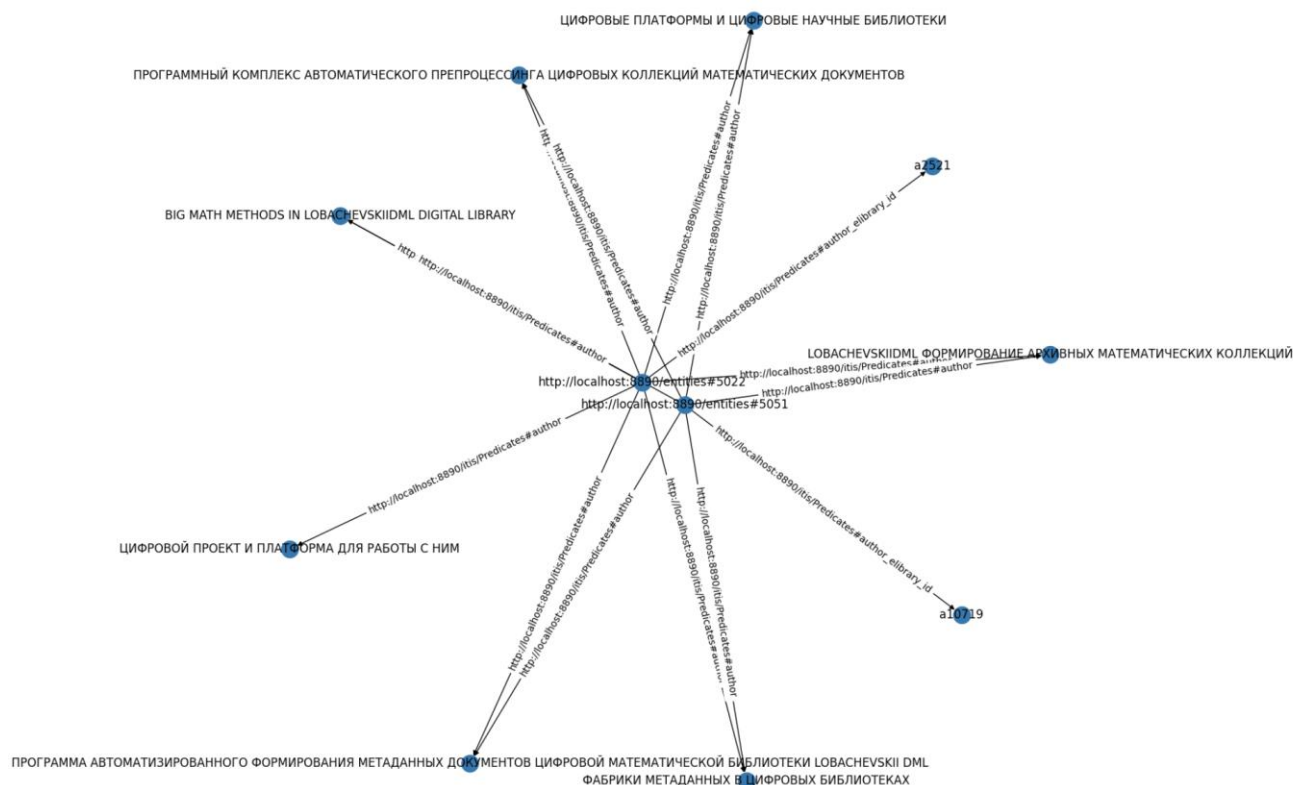


Рис. 3. Подграф графа публикаций сотрудников ИТИС

ЗАКЛЮЧЕНИЕ

Существует множество подходов для формирования цифровых пространств научных знаний и построения графов знаний. В настоящее время нет инструмента, находящегося в свободном доступе и предоставляющего возможности автоматизации наполнения графа данными и параллельного контроля их целостности. Потребность в таком инструменте имеется у исследователей, занимающихся построением графов научных знаний. Предложенный в статье метод контроля целостности данных позволил на программном уровне отслеживать тип объектов, вносимых в граф, и корректность добавляемых триплетов. Соответствующий разработанный инструмент с описанным функционалом его использования размещён в репозитории GitHub [23].

СПИСОК ЛИТЕРАТУРЫ

1. *Ataeva O., Kalenov N., Serebryakov V., Sotnikov A.* Informational Infrastructure of the Common Digital Space of Scientific Knowledge // International Conference «Common Digital Space of Scientific Knowledge», November 10–12, 2020, Moscow, Russia / CEUR Workshop Proceedings. 2021. Vol. 2990. P. 1–10.

<https://doi.org/10.51218/1613-0073-2990-1-10>

2. *Атаева О.М., Каленов Н.Е., Серебряков В.А.* Об основных понятиях Единого цифрового пространства научных знаний // Научный сервис в сети Интернет: труды XXII Всероссийской научной конференции (21–25 сентября 2020 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2020. С. 29–40.

<https://doi.org/10.20948/abrau-2020-18>

3. *Fecher B., Kahn R., Sokolovska N., Völker T., Nebe P.* Making a Research Infrastructure: Conditions and Strategies to Transform a Service into an Infrastructure // *Science and Public Policy*. 2021. Vol. 48, No. 4. P. 499–507.

<https://doi.org/10.1093/scipol/scab026>

4. *Kohn Rådberg K., Löfsten H.* Developing a knowledge ecosystem for large-scale research infrastructure // *The Journal of Technology Transfer*. 2023. Vol. 48, No. 1. P. 441–467. <https://doi.org/10.1007/s10961-022-09945-x>

5. *Papon P.* European scientific cooperation and research infrastructures: Past tendencies and future prospects // *Minerva*. 2004. Vol. 42, No. 1. P. 61–76.

<https://doi.org/10.1023/B:MINE.0000017700.63978.4a>

6. Антопольский А.Б., Каленов Н.Е., Серебряков В.А., Сотников А.Н. О едином цифровом пространстве научных знаний // Вестник Российской академии наук. 2019. Т. 89, №7. С. 728–735. <https://doi.org/10.31857/S0869-5873897728-735>

7. Elizarov A., Lipachev E. Lobachevskii Digital Library in the Scientific Space of Mathematical Knowledge, Scientific and Technical Information Processing. 2023. Vol. 50, No. 1. P. 35–39. <https://doi.org/10.3103/S0147688223010021>

8. *Serebryakov V.A., Ataeva O.M.* Ontology Based Approach to Modeling of the Subject Domain “Mathematics” in the Digital Library // *Lobachevskii Journal of Mathematics*. 2021. Vol. 42, No. 8. P. 1920–1934.

<https://doi.org/10.1134/S199508022108028X>

9. *Lange C.* Ontologies and languages for representing mathematical knowledge on the Semantic Web // *Semantic Web*. 2013. Vol. 4, No. 2. P. 119–158. <https://doi.org/10.3233/SW-2012-0059>

10. *Wang J.* Math-KG: Construction and Applications of Mathematical Knowledge Graph, arXiv:2205.03772. 2022. P. 1–5.

<https://doi.org/10.48550/arXiv.2205.03772>

11. Муромский А.А., Тучкова Н.П. Представление математических понятий в онтологии научных знаний // Онтология проектирования. 2019. Т. 9, №1 (31). С. 50–69. <https://doi.org/10.18287/2223-9537-2019-9-1-50-69>
 12. Елизаров А.М., Кириллович А.В., Липачёв Е.К., Невзорова О.А. Онтология математического знания OntoMathPRO // Доклады Российской академии наук. Математика, информатика, процессы управления. 2022. Т. 507. № 1. С. 29–35. <https://doi.org/10.31857/S2686954322700011>
 13. Елизаров А.М., Кириллович А.В., Липачёв Е.К., Невзорова О.А. Новые компоненты онтологии OntoMathPRO представления математического знания // Научный сервис в сети Интернет: труды XXV Всероссийской научной конференции (18–21 сентября 2023 г., онлайн). М.: ИПМ им. М.В. Келдыша, 2023. С. 141–151. <https://doi.org/10.20948/abrau-2023-32>
 14. Невзорова О.А., Гизатуллин Б.Т. Система автоматического построения графов знаний математических документов // Ученые записки Казанского университета. Серия: Физико-математические науки. 2023. Т. 165, № 3. С. 264–281. <https://doi.org/10.26907/2541-7746.2023.3.264-281>
 15. Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., N. Mendes P., Hellmann S., Morsey M., Van Kleef P., Auer S., Bizer C. DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia // Semantic Web. 2015. Vol. 6, No. 2. P. 167–195. <https://doi.org/10.3233/SW-140134>
 16. Vrandečić D., Krötzsch M. Wikidata: a free collaborative knowledgebase // Communications of the ACM. 2014. Vol. 57, No. 10. P. 78–85. <https://doi.org/10.1145/2629489>
 17. Schembera B., Wübbeling F., Kleikamp H., Schmidt B., Shehu A., Reidelbach M., Biedinger C., Fiedler J., Koprucki T., Iglezakis D., Göddeke D. Towards a Knowledge Graph for Models and Algorithms in Applied Mathematics // arXiv:2408.10003. <https://doi.org/10.48550/arXiv.2408.10003>
 18. Protégé. URL: <https://protege.stanford.edu>
 19. Home // RDFLib. URL: <https://rdflib.dev>
 20. SPARQL Endpoint interface to Python // SPARQLWrapper documentation URL: <https://sparqlwrapper.readthedocs.io/en/latest/main.html>
 21. RDF // Semantic Web Standards. URL: <https://www.w3.org/RDF>.
 22. Virtuoso Homepage // OpenLink Software.
-

URL: <https://virtuoso.openlinksw.com>.

23. VIGuryanov/Knowledge-Graphs-Builder // GitHub.

URL: <https://github.com/VIGuryanov/Knowledge-Graphs-Builder>.

SOFTWARE MODULE FOR FORMING DIGITAL MATHEMATICAL SPACE BASED ON KNOWLEDGE GRAPHS

V. I. Gurianov¹ [0009-0008-9126-7216], A. M. Elizarov² [0000-0003-2546-6897]

^{1, 2}Kazan Federal University, Kazan, 420008, Russia

²National Research Center (NRC) "Kurchatov Institute", Moscow, 123182, Russia

¹vadim.guryanov.2004@mail.ru, ²amelizarov@gmail.com

Abstract

The modern information space contains a lot of data, but they are often poorly structured, difficult to find and not always correct. This creates additional difficulties during researches, so digital spaces of scientific knowledge are currently being formed, in particular, based on knowledge graphs.

To ensure the quality of information, such graphs are often filled with data manually, which is time-consuming. Therefore, the creation of a tool that provides the ability to automate process of filling a graph with data, as well as ensures data quality, will simplify and speed up the process of forming digital spaces of scientific knowledge.

Methods for automating the filling of the graph with data are proposed, including parallel control of their integrity. Based on the proposed methods, a software module has been developed, the mechanisms of its functioning and its architecture are described.

Keywords: *digital space of scientific knowledge, formation of digital spaces of scientific knowledge, knowledge graphs, automation of knowledge graph construction.*

REFERENCES

1. *Ataeva O., Kalenov N., Serebryakov V., Sotnikov A.* Informational Infrastructure of the Common Digital Space of Scientific Knowledge // International Conference “Common Digital Space of Scientific Knowledge”, November 10–12, 2020, Moscow, Russia / CEUR Workshop Proceedings. 2021. Vol. 2990. P. 1–10.
<https://doi.org/10.51218/1613-0073-2990-1-10>
2. *Ataeva O.M., Kalenov N.E., Serebryakov V.A.* On the Basic Concepts of the Unified Digital Space of Scientific Knowledge // Scientific service on the Internet: Proceedings of the XXII All-Russian scientific conference (September 21–25, 2020, online). Moscow: Keldysh Institute of Applied Mathematics, 2020. P. 29–40.
<https://doi.org/10.20948/abrau-2020-18>
3. *Fecher B., Kahn R., Sokolovska N., Völker T., Nebe P.* Making a Research Infrastructure: Conditions and Strategies to Transform a Service into an Infrastructure // Science and Public Policy. 2021. Vol. 48, No. 4. P. 499–507.
<https://doi.org/10.1093/scipol/scab026>
4. *Kohn Rådberg K., Löfsten H.* Developing a knowledge ecosystem for large-scale research infrastructure // The Journal of Technology Transfer. 2023. Vol. 48, No. 1. P. 441–467. <https://doi.org/10.1007/s10961-022-09945-x>
5. *Papon P.* European scientific cooperation and research infrastructures: Past tendencies and future prospects // Minerva. 2004. Vol. 42, No. 1. P. 61–76.
<https://doi.org/10.1023/B:MINE.0000017700.63978.4a>
6. *Antopolsky A.B., Kalenov N.E., Serebryakov V.A., Sotnikov A.N.* On a single digital space of scientific knowledge // Bulletin of the Russian Academy of Sciences. 2019. Vol. 89, No. 7. P. 728–735. <https://doi.org/10.31857/S0869-5873897728-735>
7. *Elizarov A., Lipachev E.* Lobachevskii Digital Library in the Scientific Space of Mathematical Knowledge, Scientific and Technical Information Processing. 2023. Vol. 50, No. 1. P. 35–39. <https://doi.org/10.3103/S0147688223010021>
8. *Serebryakov V.A., Ataeva O.M.* Ontology Based Approach to Modeling of the Subject Domain “Mathematics” in the Digital Library // Lobachevskii Journal of Mathematics. 2021. Vol. 42, No. 8. P. 1920–1934.
<https://doi.org/10.1134/S199508022108028X>

9. Lange C. Ontologies and languages for representing mathematical knowledge on the Semantic Web // *Semantic Web*. 2013. Vol. 4, No. 2. P. 119–158. <https://doi.org/10.3233/SW-2012-0059>
10. Wang J. Math-KG: Construction and Applications of Mathematical Knowledge Graph, arXiv:2205.03772. 2022. P. 1–5. <https://doi.org/10.48550/arXiv.2205.03772>
11. Muromskiy A.A., Tuchkova N.P. Representation of mathematical concepts in the ontology of scientific knowledge // *Ontology of Design*. 2019. Vol. 9, No. 1 (31). P. 50–69. <https://doi.org/10.18287/2223-9537-2019-9-1-50-69>
12. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A. Ontology of mathematical knowledge OntoMathPRO // *Reports of the Russian Academy of Sciences. Mathematics, informatics, control processes*. 2022. Vol. 507, No. 1. P. 29–35. <https://doi.org/10.31857/S2686954322700011>
13. Elizarov A.M., Kirillovich A.V., Lipachev E.K., Nevzorova O.A. New components of the OntoMathPRO ontology for representing mathematical knowledge // *Scientific service on the Internet: proceedings of the XXV All-Russian Scientific Conference (September 18–21, 2023, online)*. Moscow: Keldysh Institute of Applied Mathematics, 2023. P. 141–151. <https://doi.org/10.20948/abrau-2023-32>
14. Nevzorova O.A., Gizatullin B.T. System for automatic construction of knowledge graphs of mathematical documents // *Scientific Notes of Kazan University. Series: Physical and Mathematical Sciences*. 2023. Vol. 165, No. 3. P. 264–281. <https://doi.org/10.26907/2541-7746.2023.3.264-281>
15. Lehmann J., Isele R., Jakob M., Jentzsch A., Kontokostas D., N. Mendes P., Hellmann S., Morsey M., Van Kleef P., Auer S., Bizer C. DBpedia – A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia // *Semantic Web*. 2015. Vol. 6, No. 2. P. 167–195. <https://doi.org/10.3233/SW-140134>
16. Vrandečić D., Krötzsch M. Wikidata: a free collaborative knowledgebase // *Communications of the ACM*. 2014. Vol. 57, No. 10. P. 78–85. <https://doi.org/10.1145/2629489>
17. Schembera B., Wübbeling F., Kleikamp H., Schmidt B., Shehu A., Reidelbach M., Biedinger C., Fiedler J., Koprucki T., Iglezakis D., Göddeke D. Towards a Knowledge Graph for Models and Algorithms in Applied Mathematics // arXiv:2408.10003.

<https://doi.org/10.48550/arXiv.2408.10003>

18. Protégé. URL: <https://protege.stanford.edu>

19. Home // RDFLib. URL: <https://rdflib.dev>

20. SPARQL Endpoint interface to Python // SPARQLWrapper documentation

URL: <https://sparqlwrapper.readthedocs.io/en/latest/main.html>

21. RDF // Semantic Web Standards. URL: <https://www.w3.org/RDF>.

22. Virtuoso Homepage // OpenLink Software.

URL: <https://virtuoso.openlinksw.com>

23. VIGuryanov/Knowledge-Graphs-BUILDER // GitHub.

URL: <https://github.com/VIGuryanov/Knowledge-Graphs-BUILDER>

СВЕДЕНИЯ ОБ АВТОРАХ



ГУРЬЯНОВ Вадим Игоревич – студент Института информационных технологий и интеллектуальных систем Казанского федерального университета.

Научные интересы: единое пространство научных математических знаний, автоматизация построения графов знаний, технологии извлечения знаний.

Vadim Igorevich GURIANOV – student of the Institute of Information Technologies and Intelligent Systems of Kazan Federal University.

Research interests: common space of scientific mathematical knowledge, automation of knowledge graph construction, knowledge extraction technologies.

email: vadim.guryanov.2004@mail.ru

ORCID: 0009-0008-9126-7216



ЕЛИЗАРОВ Александр Михайлович – доктор физико-математических наук, профессор, заслуженный деятель науки Республики Татарстан и Российской Федерации, профессор кафедры программной инженерии Института информационных технологий и интеллектуальных систем Казанского федерального университета.

Научные интересы: цифровые библиотеки, единое пространство научных математических знаний, интеллектуальный анализ данных, рекомендательные системы, облачные вычисления, технологии извлечения знаний.

Alexander Mikhailovich ELIZAROV – Doctor of Physics and Mathematics, Professor, Honored Scientist of the Republic of Tatarstan and the Russian

Federation, Professor of the Department of Software Engineering of the Institute of Information Technologies and Intelligent Systems of Kazan Federal University.

Research interests: digital libraries, common space of scientific mathematical knowledge, data mining, recommender systems, cloud computing, knowledge extraction technologies.

email: amelizarov@gmail.com

ORCID: 0000-0003-2546-6897

Материал поступил в редакцию 5 февраля 2025 года

УДК 004.432+519.673

ОПТИМИЗАЦИЯ АЛГОРИТМОВ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ C++ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МНОГОПОТОЧНОСТИ

Ю. С. Ефимов^[0009-0003-6333-2109]

H&H company, г. Лиссабон, 1675-107, Португалия

yura2x2@gmail.com

Аннотация

Представлены основные методы численного моделирования (конечных разностей, конечных элементов, Монте-Карло, Рунге–Кутты). Рассмотрены основные параметры, используемые для оптимизации алгоритмов численного моделирования с точки зрения длительности выполнения кода и эффективного использования ресурсов процессора. Проанализированы основные недостатки многопоточности, связанные с синхронизацией данных, дедлоками и состояниями гонки и методы их устранения на основе применения мьютексов и атомарных операций на примере метода Монте-Карло.

Ключевые слова: язык программирования C++, методы многопоточности, численное моделирование, синхронизация данных.

ВВЕДЕНИЕ

Численное моделирование играет ключевую роль в современных научных исследованиях и технических разработках [1]. Оно позволяет анализировать сложные физические, биологические и экономические процессы с высокой точностью [2]. Однако вычислительная сложность многих численных методов требует эффективного использования аппаратных ресурсов. Одним из наиболее продуктивных подходов к ускорению вычислений является многопоточность в языке C++ [3], которая позволяет распределять задачи между несколькими потоками на многоядерных процессорах.

Многопоточность в C++ предоставляет мощные инструменты для параллелизации вычислительных алгоритмов. Используя библиотеки, такие как `std::thread`, и механизмы синхронизации, такие как мьютексы и условные пере-

менные, разработчики могут разделить задачу на независимые подзадачи и выполнять их параллельно. Это особенно эффективно для численных методов, где операции над большими массивами данных или сложные итерационные процессы могут быть разделены между несколькими ядрами процессора.

Применение многопоточности в численном моделировании требует тщательного анализа алгоритма и выбора оптимальной стратегии параллелизации. Важно учитывать накладные расходы на создание и синхронизацию потоков, а также потенциальные проблемы, связанные с гонками данных и взаимными блокировками. Эффективная реализация многопоточности позволяет существенно сократить время вычислений и повысить производительность численных моделей [4].

Примерами успешного применения многопоточности при численном моделировании являются расчеты в гидродинамике, моделирование климата, анализ данных в физике высоких энергий и оптимизация логистических процессов. В этих областях многопоточные алгоритмы позволяют проводить более масштабные и точные симуляции, что приводит к новым научным открытиям и технологическим инновациям.

Таким образом, многопоточность в C++ является важным инструментом для повышения эффективности численного моделирования. Правильное применение этой технологии позволяет значительно ускорить вычисления и расширить возможности анализа сложных процессов, открывая новые горизонты в научных исследованиях и инженерных разработках.

Целью настоящей работы является исследование возможностей многопоточного программирования в C++ для оптимизации алгоритмов численного моделирования. Рассмотрены основные принципы многопоточного программирования, методы балансировки нагрузки и синхронизации потоков, а также практические примеры реализации многопоточных алгоритмов.

ОСНОВЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Численное моделирование включает в себя использование математических методов для приближенного решения уравнений, описывающих сложные процессы. Оно применяется в таких областях, как аэродинамика, климатология, финансовый анализ, медицинская диагностика и др.

Наиболее популярные численные методы включают (см., например, [4]):

1. **Метод конечных разностей** – это численный способ решения дифференциальных уравнений, где производные аппроксимируются разностными отношениями. Он относится к классу сеточных методов.

2. **Численный метод конечных элементов (МКЭ)** представляет собой вычислительную стратегию для нахождения решений дифференциальных уравнений в частных производных и интегральных уравнений, которые часто возникают в различных областях прикладных математики и физики. Он широко используется для исследования деформаций в твердых телах, процессов теплопередачи, гидродинамических явлений, электромагнитных полей и задач оптимизации формы.

3. **Методы Монте-Карло (ММК)** – это совокупность вычислительных приемов, предназначенных для анализа случайных событий. Ключевой принцип заключается в имитации процесса с использованием генератора случайных чисел, многократном выполнении моделирования и последующем расчете вероятностных характеристик исследуемого процесса на основе полученных результатов. В качестве примера, для определения среднего расстояния между произвольными точками в круге создают случайные пары координат, вычисляют расстояние для каждой пары, а затем определяют среднее арифметическое значение.

4. **Методы Рунге–Кутты** – это обширная группа численных алгоритмов, предназначенных для решения задачи Коши для обыкновенных дифференциальных уравнений и их систем.

ОСНОВЫ МНОГОПОТОЧНОСТИ В C++

Многопоточность позволяет программе одновременно выполнять несколько задач, используя ресурсы многоядерных процессоров. В C++ многопоточное программирование реализуется с помощью библиотеки `<thread>`.

Пример простого многопоточного кода в C++:

```
#include <iostream>
#include <thread>
void task() {
    std::cout << "Поток выполняется" << std::endl;
}
```

```
int main() {  
    std::thread t(task); // Создание потока  
    t.join(); // Ожидание завершения потока  
    return 0;  
}
```

Метод `.join ()` гарантирует, что основной поток дождётся завершения порожденного потока.

При использовании многопоточности возникают сложности, связанные с:

- 1) **синхронизацией данных** – необходимость защиты общих ресурсов от одновременного изменения;
- 2) **состояниями гонки** (race conditions) – ситуациями, когда несколько потоков изменяют переменные одновременно;
- 3) **дедлоками** (deadlocks) – взаимоблокировками, приводящими к зависанию программы.

Для решения этих проблем используют механизмы синхронизации, такие как мьютексы (`std::mutex`) и атомарные операции (`std::atomic`).

Пример использования мьютекса:

```
#include <iostream>  
#include <thread>  
#include <mutex>  
std::mutex mtx;  
  
void printData(int id) {  
    std::lock_guard<std::mutex> lock(mtx);  
    std::cout << "Поток " << id << " выполняется\n";  
}  
  
int main() {  
    std::thread t1(printData, 1);  
    std::thread t2(printData, 2);  
  
    t1.join();  
    t2.join();  
  
    return 0;  
}
```

Использование `std::lock_guard` предотвращает дедлоки, автоматически освобождая мьютекс при выходе из области видимости.

ОПТИМИЗАЦИЯ МНОГОПОТОЧНЫХ ВЫЧИСЛЕНИЙ

Правильное распределение вычислений между потоками позволяет избежать простаивания процессорных ядер и повысить эффективность выполнения программы. Использование динамического распределения задач и алгоритмов балансировки нагрузки играет важную роль в оптимизации.

Переключение между потоками требует значительных затрат ресурсов. Использование эффективных структур данных, уменьшение количества синхронизаций и минимизация блокировок позволяют снизить накладные расходы.

ПРАКТИЧЕСКОЕ ПРИМЕНЕНИЕ МНОГОПОТОЧНОСТИ В ЧИСЛЕННОМ МОДЕЛИРОВАНИИ

В качестве типичного примера вычислительного моделирования можно привести метод Монте-Карло. Он широко применяется для решения задач интегрирования, проведения симуляций в физических исследованиях и построения моделей для анализа финансовых рынков.

В качестве иллюстрации рассмотрим многопоточную реализацию метода Монте-Карло для приближенного определения значения числа π :

```
#include <iostream>
#include <thread>
#include <vector>
#include <random>
#include <atomic>

const int num_threads = 4;
const int num_points = 1000000;
std::atomic<int> inside_circle(0);
void monteCarlo(int num_samples) {
    std::random_device rd;
    std::mt19937 gen(rd());
    std::uniform_real_distribution<double> dist(-1.0, 1.0);

    int count = 0;
    for (int i = 0; i < num_samples; i++) {
```

```
double x = dist(gen);
double y = dist(gen);
if (x * x + y * y <= 1.0) count++;
}
inside_circle += count;
}

int main() {
    std::vector<std::thread> threads;

    for (int i = 0; i < num_threads; i++) {
        threads.emplace_back(monteCarlo, num_points / num_threads);
    }
    for (auto& t : threads) {
        t.join();
    }
    double pi = 4.0 * inside_circle / num_points;
    std::cout << "Приближенное значение числа π: " << pi << std::endl;
    return 0;
}
```

В данном коде каждый поток независимо выполняет расчет, а переменная `inside_circle` обновляется атомарно для предотвращения состояния гонки.

Библиотека OpenMP позволяет упростить распараллеливание циклов без явного управления потоками.

Пример параллельного вычисления суммы массива:

```
#include <iostream>
#include <omp.h>
int main() {
    const int N = 100000;
    int arr[N], sum = 0;
    for (int i = 0; i < N; i++) arr[i] = i;
    #pragma omp parallel for reduction(+:sum)
    for (int i = 0; i < N; i++) {
        sum += arr[i];
    }
    std::cout << "Сумма элементов массива: " << sum << std::endl;
    return 0;
}
```



```
}
```

Использование `#pragma omp parallel for` автоматически распределяет итерации цикла между потоками, а `reduction(+:sum)` гарантирует корректное суммирование.

ПРОИЗВОДИТЕЛЬНОСТЬ И ТЕСТИРОВАНИЕ

Коэффициент ускорения многопоточной программы вычисляется по формуле

$$S = T_1/T_n,$$

где T_1 – время выполнения на одном потоке, а T_n – время выполнения на нескольких потоках.

Для измерения времени выполнения можно использовать `std::chrono`:

```
#include <iostream>
```

```
#include <chrono>
```

```
int main() {
```

```
    auto start = std::chrono::high_resolution_clock::now();
```

```
    // Ваш код
```

```
    auto end = std::chrono::high_resolution_clock::now();
```

```
    std::chrono::duration<double> elapsed = end - start;
```

```
    std::cout << "Время выполнения: " << elapsed.count() << " секунд" << std::endl;
```

```
    return 0;
```

```
}
```

Необходимо равномерно распределять вычисления между потоками, чтобы избежать ситуаций, когда один поток загружен больше других. Для этого можно использовать динамическое распределение в OpenMP:

```
#pragma omp parallel for schedule(dynamic)
```

ГЛУБОКАЯ ОПТИМИЗАЦИЯ МНОГОПОТОЧНЫХ АЛГОРИТМОВ

При разработке многопоточных алгоритмов важно выбирать структуры данных, которые минимизируют блокировки потоков. Некоторые из наиболее эффективных подходов таковы:

- 1) **атомарные структуры данных** (`std::atomic`, `std::atomic_flag`) – используются для защиты переменных от одновременной записи. Они эффективнее мьютексов, так как работают без блокировки;
- 2) **параллельные контейнеры** (например, `concurrent_queue` в TBB) – обеспечивают потокобезопасное взаимодействие без необходимости вручную управлять мьютексами;
- 3) **чтение – копирование – обновление** (Read–Copy–Update, RCU) – техника, позволяющая избежать блокировок при работе с разделяемыми структурами.

Пример использования `std::atomic` в многопоточном счетчике:

```
#include <iostream>
#include <thread>
#include <atomic>
std::atomic<int> counter(0);
void increment(int iterations) {
    for (int i = 0; i < iterations; i++) {
        counter.fetch_add(1, std::memory_order_relaxed);
    }
}
int main() {
    const int num_threads = 4;
    const int iterations_per_thread = 1000000;
    std::thread threads[num_threads];
    for (int i = 0; i < num_threads; i++) {
        threads[i] = std::thread(increment, iterations_per_thread);
    }
    for (auto& t : threads) {
        t.join();
    }
    std::cout << "Итоговое значение счетчика: " << counter.load() << std::endl;
```

```
    return 0;  
}
```

Этот метод позволяет избежать проблем состояния гонки, улучшая стабильность программы.

Производительность многопоточных программ может снижаться из-за ложного разделения – ситуации, когда разные потоки работают с разными переменными, но эти переменные находятся в одной кэш-линии процессора.

Методы устранения ложного разделения:

- 1) выравнивание данных по размеру кэш-линии (обычно 64 байта);
- 2) использование `alignas(64)` или `std::hardware_destructive_interference_size`;
- 3) разделение потоков по разным частям памяти.

Пример устранения ложного разделения с `alignas(64)`:

```
#include <iostream>  
#include <thread>  
#include <vector>  
  
const int num_threads = 4;  
alignas(64) int data[num_threads] = {0};  
void increment(int index) {  
    for (int i = 0; i < 1000000; i++) {  
        data[index]++;  
    }  
}  
  
int main() {  
    std::vector<std::thread> threads;  
    for (int i = 0; i < num_threads; i++) {  
        threads.emplace_back(increment, i);  
    }  
    for (auto& t : threads) {  
        t.join();  
    }  
    for (int i = 0; i < num_threads; i++) {  
        std::cout << "data[" << i << "] = " << data[i] << std::endl;  
    }  
}
```

```
    return 0;
}
```

Вместо явного управления потоками с `std::thread` можно использовать `std::async`, который позволяет запускать задачи асинхронно с автоматическим управлением ресурсами.

Пример использования `std::async`:

```
#include <iostream>
#include <future>

int slowCalculation() {
    std::this_thread::sleep_for(std::chrono::seconds(2));
    return 42;
}

int main() {
    std::future<int> result = std::async(std::launch::async, slowCalculation);
    std::cout << "Ждем результата..." << std::endl;
    std::cout << "Результат: " << result.get() << std::endl;
    return 0;
}
```

Применение `std::async` упрощает структуру программы по сравнению с использованием `std::thread` в сочетании с `std::mutex`, поскольку механизм синхронизации в этом случае становится прозрачным и автоматизированным.

ПРОИЗВОДИТЕЛЬНОСТЬ И ТЕСТИРОВАНИЕ МНОГОПОТОЧНЫХ АЛГОРИТМОВ

Для анализа производительности параллельных вычислений:

- 1) время выполнения кода оценивается с помощью библиотек `std::chrono` для точного измерения временных интервалов;
- 2) для оценки загрузки процессора применяются такие инструменты, как системные команды `top`, `htop` и специализированное приложение `perf` в операционных системах на базе Linux;
- 3) ключевыми инструментами профилирования являются `gprof` для Unix-подобных систем, `Intel VTune` и профайлер Visual Studio для оценки эффективности кода.

Пример измерения времени выполнения программы:

```
#include <iostream>
#include <chrono>
#include <thread>
void task() {
    std::this_thread::sleep_for(std::chrono::seconds(1));
}
int main() {
    auto start = std::chrono::high_resolution_clock::now();
    std::thread t(task);
    t.join();
    auto end = std::chrono::high_resolution_clock::now();
    std::chrono::duration<double> elapsed = end - start;
    std::cout << "Время выполнения: " << elapsed.count() << " секунд" << std::endl;
    return 0;
}
```

Коэффициент ускорения (speedup) рассчитывается по формуле

$$S = T_1/T_p,$$

где T_1 – время выполнения на одном потоке, а T_p – время выполнения на p потоках.

Коэффициент эффективности (efficiency) определяется как

$$E = S/p.$$

При идеальном масштабировании S приближается к p , а E – к 1.

ПЕРСПЕКТИВЫ РАЗВИТИЯ МНОГОПОТОЧНОГО ПРОГРАММИРОВАНИЯ В ЧИСЛЕННОМ МОДЕЛИРОВАНИИ

Рассмотрим использование CUDA (NVIDIA) и OpenCL для переноса вычислений на видеокарту.

Параллелизация численных методов на GPU дает прирост производительности в 10–100 раз.

Пример использования CUDA:

```
__global__ void add(int *a, int *b, int *c, int N) {
    int i = blockIdx.x * blockDim.x + threadIdx.x;
```

```
if (i < N) c[i] = a[i] + b[i];  
}
```

Многие процессоры поддерживают инструкции SIMD, позволяющие выполнять несколько операций одновременно.

Пример компиляции с автопараллелизацией в GCC:

```
g++ -O2 -march=native -ftree-vectorize program.cpp -o program
```

ПЕРСПЕКТИВЫ РАЗВИТИЯ

1. Использование векторизации с SIMD-инструкциями для повышения эффективности вычислений.
2. Применение GPU-ускорения с помощью CUDA и OpenCL для численного моделирования.
3. Развитие асинхронного программирования с использованием `std::async` и `future`.

ЗАКЛЮЧЕНИЕ

Использование многопоточности в C++ позволяет значительно повысить производительность численного моделирования, снижая время выполнения ресурсоемких расчетов. Однако для эффективного применения многопоточных технологий важно учитывать балансировку нагрузки, минимизацию состояния гонки и снижение накладных расходов на управление потоками.

СПИСОК ЛИТЕРАТУРЫ

1. Умнов А.Е. Методы математического моделирования: учебное пособие. М.: МФТИ, 2013. 295 с.
2. Марчевский И.К., Щерица О.В. Численные методы решения задач математической физики: учебно-методическое пособие. М.: МГТУ им. Н.Э. Баумана, 2016. 64 с.
3. Уильям Э. Практика многопоточного программирования C++. Параллельное программирование на C++ в действии. Практика разработки многопоточных программ. Пер. с англ. Слинкин А.А. М.: ДМК Пресс, 2012. 672с.
4. Зенков А.В. Численные методы: учеб. пособие. Екатеринбург: Изд-во Урал. ун-та, 2016. 124 с.

OPTIMIZATION OF NUMERICAL SIMULATION ALGORITHMS C++ WITH MULTITHREADING METHODS

Yu. S. Efimov^[0009-0003-6333-2109]

H&H company, Lisbon, 1675-156, Portugal

yura2x2@gmail.com

Abstract

The main methods of numerical simulation (finite difference method, finite element method, Monte Carlo method, Runge–Kutta method) are presented. The main parameters used to optimize numerical modeling algorithms in terms of code execution time and efficient use of processor resources are considered. The main disadvantages of multithreading related to data synchronization, deadlocks and race conditions and methods for eliminating them based on the use of mutexes and atomic operations using the Monte Carlo method as an example were analyzed.

Keywords: *programming language C++, multithreading methods, numerical simulation, data synchronization.*

REFERENCES

1. *Umnov A.E. Metody matematicheskogo modelirovaniya: uchebnoe posobie.* M.: MFTI, 2013. 295 s. (in Russian).
2. *Marchevskij I.K., Shcherica O.V. Chislennye metody resheniya zadach matematicheskoy fiziki: uchebno-metodicheskoe posobie.* M.: MGTU im. N.E. Baumana, 2016. 64 s. (in Russian).
3. *Uil'yam E. Praktika mnogopotochnogo programirovaniya S++. Parallel'noe programirovanie na C++ v dejstvii. Praktika razrabotki mnogopotochnyh programm.* Per. s angl. Slinkin A.A. M.: DMK Press, 2012. 672s. (in Russian).
4. *Zenkov A.V. Chislennye metody: ucheb. posobie.* Ekaterinburg: Izd-vo Ural. un-ta, 2016. 124 s. (in Russian).

СВЕДЕНИЯ ОБ АВТОРЕ



ЕФИМОВ Юрий Сергеевич – 1979 года рождения, учился в Санкт-Петербургском государственном университете аэрокосмического приборостроения по специальности 2202 автоматизированные системы обработки информации и управления. В настоящее время работает программистом C++ в компании H&H. Область научных интересов: параллельные и распределённые вычисления, системы с низкой задержкой, оптимизация производительности программного обеспечения.

Yuri Sergeevich EFIMOV – born in 1979, studied at Saint Petersburg State University of Aerospace Instrumentation, specializing in Automated Information Processing and Control Systems (specialty code 2202). He currently works as a C++ programmer at H&H. Research interests: parallel and distributed computing, low-latency systems, software performance optimization.

email: yura2x2@gmail.com

ORCID: 0009-0003-6333-2109

Материал поступил в редакцию 2 апреля 2025 года

МЕТОДИКА СРАВНЕНИЯ ПРОГРАММНЫХ РЕШЕНИЙ РАСПОЗНАВАНИЯ ТЕКСТОВ НАУЧНЫХ ПУБЛИКАЦИЙ ПО КАЧЕСТВУ ИЗВЛЕЧЕНИЯ МЕТАДААННЫХ

И. И. Кузнецов¹ [0009-0001-6287-8295], О. П. Новиков² [0009-0009-3494-3799],
Д. Ю. Ильин³ [0000-0002-0241-2733]

^{1, 2}Российский государственный университет им. А.Н. Косыгина (Технологии.
Дизайн. Искусство), г. Москва, 119071, Россия

³МИРЭА – Российский технологический университет, г. Москва, 119454, Россия

¹iliya-kuznetsov@mail.ru, ²novikovop55@rambler.ru, ³i@dmitryilin.com

Аннотация

Метаданные научных публикаций используются для построения каталогов, определения цитируемости публикаций и решения других задач. Автоматизация извлечения метаданных из PDF-файлов позволяет ускорить выполнение обозначенных задач, а от качества извлеченных данных зависит возможность их дальнейшего использования. Проанализированы существующие программные решения, в итоге отобраны три: GROBID, CERMINE, ScientificPdfParser. Предложена методика сравнения этих программных решений распознавания текстов научных публикаций по качеству извлечения метаданных. На основе методики проведен эксперимент по извлечению четырех типов метаданных (название, аннотация, дата публикации, имена авторов). Для сравнения программных решений использован набор из 112457 публикаций с разбиением на 23 предметные области, сформированный на основе данных Semantic Scholar. Приведен пример выбора эффективного программного решения извлечения метаданных в условиях заданных приоритетов для предметных областей и типов метаданных с использованием взвешенной суммы. Определено, что для приведенного примера CERMINE показывает эффективность на 10,5% выше, чем GROBID, и на 9,6% выше, чем ScientificPdfParser.

Ключевые слова: распознавание текста, научные публикации, метаданные, качество извлечения данных, методика.

ВВЕДЕНИЕ

Научные статьи являются одним из основных способов публикации и предоставления результатов разнообразных научных исследований. Помимо непосредственно информации об исследованиях, содержащейся в текстах статей, интерес представляют и относящиеся к этим статьям метаданные. К таким метаданным относятся названия статей, сведения об авторах, дате и месте публикации, используемом цитировании и др. Сбор и анализ подобных метаданных может служить различным целям – рассмотрению связей статей на уровне цитирования [1], анализу научной области на предмет того, насколько активно ведется публикация связанных с ней работ в какой-то период времени и какие направления исследований в ней преобладают [2], формирования выборки статей по авторам и другим признакам, и многое другое.

Сбор метаданных, а также какого-либо контента статей вручную является трудоемким. Электронные библиотеки, содержащие большие объемы научных статей, помимо текстов статей могут предоставлять и связанные с ними метаданные, однако на практике далеко не все источники статей содержат их метаданные или же предоставляют лишь часть из возможных данных. Кроме того, такие метаданные могут быть низкого качества: содержать ошибки, неточности, пустые поля и т. д. [3]. Таким образом, актуальным становится извлечение метаданных и контента из статей в автоматизированном режиме, и для этого существует ряд программных решений.

Исследования, связанные с автоматизированным извлечением метаданных и конкретных фрагментов статей, ведутся в различных направлениях. В ряде работ, например [4–6], предложены оригинальные модели и программные решения для извлечения метаданных. При этом одни решения специализируются на извлечении библиографических ссылок, такие как модель полного цикла извлечения и постобработки [7, 8], другие – на извлечении непосредственно метаданных статьи [9]. Некоторые работы посвящены извлечению формул и результатов измерений из научных статей: в [10] описана такая система, разработанная на основании GROBID, в [11] – система, использующаяся для извлечения из текстов статей информации о различных материалах (названий, формул, свойств).

Аналогичные разработки имеются для извлечения таких данных из научных публикаций в области сверхпроводников [12] или, например, в геологии [13].

Программные решения, изначально созданные для работы с научными статьями, находят применение и при извлечении данных из других документов сходной структуры: так, в [14] описано использование инструмента на основе GROBID для извлечения результатов измерений из медицинских отчетов. Помимо разработки новых моделей для извлечения метаданных или же извлечения таких фрагментов, как результаты измерений, некоторые исследователи предлагают методы использования результатов измерений для оценки качества самих статей или их оформления. В [15] извлечение названия, ключевых слов и полного набора библиографических ссылок использовано для проверки того, насколько содержание статьи и тематики цитируемых работ соответствуют заявленной тематике исходной статьи. В работе [16] извлечение ключевых слов и фрагментов статьи из документа использовано для оценки уровня репрезентативности аннотации относительно содержания статьи, а в [17] распознавание структуры и извлечение элементов текста применены для создания аннотаций и кратких описаний статей, в которых представлены основные тезисы исследования.

Существующие решения, как правило, работают с одним языком, например английским, ввиду чего еще одним направлением является разработка решений для поддержки мультязычных наборов статей. Например, в [18] была предложена модель извлечения библиографических ссылок, поддерживающая многие языки (включая корейский), а в работах [19] и [20] авторы сосредоточились на создании размеченных наборов данных для поддержки статей, написанных на кириллице.

Поскольку существуют готовые программные решения для извлечения метаданных, ссылок и контента статей, проводятся и исследования (например, [21, 22]) по сравнению их эффективности и качества извлечения. В [21] авторы провели сравнение трех инструментов для извлечения библиографических ссылок на основе статей из индонезийской базы научных журналов и предположили использовать эти результаты при создании приложения для корректного распознавания и извлечения библиографических ссылок. В работе [22] авторами пред-

ложена инфраструктура для измерения качества и производительности извлечения метаданных при сравнении качества извлечения несколькими различными программными решениями.

При извлечении метаданных могут быть допущены различные ошибки: неверное распознавание элементов текста и ошибки разметки, ошибки при извлечении конкретных символов или их неверная интерпретация, некорректное распознавание и извлечение строки текста и др. Большое количество ошибок будет означать низкое качество извлечения метаданных, а полученные данные будет затруднительно или невозможно использовать для последующих анализа и обработки. Поскольку извлечение метаданных может осуществляться на начальных этапах в цепочке анализа данных, от качества извлечения будет зависеть корректность результатов всей последующей части цепочки.

Целью настоящей работы является разработка методики сравнения программных решений распознавания текстов научных публикаций по качеству извлечения метаданных. Отличиями от ранее представленных работ являются используемая процедура определения пороговых значений метрик схожести строк и учет влияния принадлежности публикаций к различным предметным областям на качество извлечения метаданных. При этом под качеством понимаются точность и полнота извлечения метаданных. Точность отражает верное распознавание всех символов, а также отсутствие ошибок и искажений при извлечении строк и фрагментов, содержащих метаданные. Полнота же является показателем того, что все метаданные указанного типа будут извлечены из документа (например, будут извлечены все авторы публикации).

1. АНАЛИЗ СУЩЕСТВУЮЩИХ ПРОГРАММНЫХ РЕШЕНИЙ

Существует ряд программных решений, позволяющих извлекать метаданные из научных статей в формате PDF. Они различаются по набору извлекаемых метаданных, уровню поддержки, затрачиваемым ресурсам и т. д. При этом алгоритм их работы состоит, как правило, из набора этапов, схожих по принципу организации. На первом этапе происходят распознавание символов и реконструкция базовой структуры документа: каждой странице ставится в соответствие набор текстовых блоков. На втором этапе текстовые блоки классифицируются, происходит определение принадлежности блоков тому или иному типу элементов статьи.

На третьем этапе из распознанных элементов извлекаются соответствующие им метаданные (а также текстовый контент, изображения, таблицы и т. д. в зависимости от возможностей программы). В совокупности, на втором и третьем этапах для классификации блоков и распознавания и извлечения метаданных и контента используются различные методы: подходы на основе правил [23], эвристических предположений [24], различных алгоритмов машинного обучения [24–26] или же комбинации этих методов [27].

Рассмотрены существующие решения для извлечения метаданных и контента из научных статей, а именно: CERMINE [28], GROBID [29], ParsCit [30], Neural-ParsCit [31], PDFX [32], ScientificPdfParser, Science Parse и SPv2. При распознавании и извлечении метаданных в этих решениях используются различные методы. В CERMINE – это метод опорных векторов и алгоритм кластеризации k-средних для извлечения ссылок. В GROBID алгоритм основан на методе условных случайных полей. Этот же метод используется в ParsCit в сочетании с эвристическими методами на основе регулярных выражений. Решение Neural-ParsCit, являющееся усовершенствованной версией ParsCit, использует модель глубокого обучения LSTM. ScientificPdfParser имеет три режима извлечения: на основании заранее определенных правил и эвристики, с использованием машинного обучения на основе наивного байесовского алгоритма, а также гибридный режим с генерацией обучающей модели. Science Parse и SPv2 тоже используют методы машинного обучения, включая метод условных случайных полей. В табл. 1 приведены некоторые характеристики перечисленных решений, которые рассматривались при их отборе для эксперимента.

Таблица 1. Характеристики программных решений извлечения метаданных научных статей

Признак	CERMINE	GROBID	ParsCit/Neural-ParsCit	PDFX	ScientificPdfParser	Science Parse	SPv2
Язык программы	Java	Java	Perl	Java	Java	Java + Scala	Java + Scala
Открытый исходный	Да	Да	Да	Нет	Да	Да	Да

код							
Извлечение метаданных	Да	Да	Нет	Да	Да	Да	Да
Извлечение ссылок	Да	Да	Да	Да	Да	Да	Да
Извлечение метаданных для ссылок	Да	Да	Да	Нет	Да	Да	Нет
Количество типов метаданных	11	7	-	4	5	4	3
Тип выходного файла	XML	JSON	XML	XML	JSON	XML	XML

После рассмотрения этих решений для эксперимента были отобраны несколько из них. Условиями для отбора служили: наличие открытого исходного кода и возможность извлекать метаданные статьи (не только библиографические ссылки). Согласно этим условиям были исключены PDFX как не обладающий открытым исходным кодом и обе версии ParsCit, поскольку они извлекают только библиографические данные. Кроме того, из рассмотрения были исключены Science Parse и SPv2, так как из-за долгого отсутствия поддержки авторами кодовой базы не удалось достичь их корректной и стабильной работы без дополнительной актуализации и доработки кода. Таким образом, для эксперимента были взяты три решения: CERMINE, GROBID и ScientificPdfParser.

2. МЕТОДИКА СРАВНЕНИЯ ПРОГРАММНЫХ РЕШЕНИЙ ПО КАЧЕСТВУ ИЗВЛЕЧЕНИЯ МЕТАДААННЫХ

Для эксперимента должны использоваться научные статьи в формате PDF, собранные на основании какой-либо электронной библиотеки таких статей. Эта же библиотека должна предоставлять метаданные для выбранных статей, что послужит контрольным набором для сравнения с результатами извлечения. Кроме того, должны быть отобраны статьи из различных научных направлений, что позволит получить как общие оценки качества извлечения, так и оценки для различ-

ных направлений, поскольку принципы оформления и структуры статей могут отличаться от направления к направлению. Для проведения эксперимента по сравнению программных решений по качеству извлечения метаданных из научных статей была разработана методика, содержащая следующие шаги.

1. Выбор подходящей электронной библиотеки, содержащей метаданные для статей различных научных направлений и PDF-файлы с текстами научных публикаций либо ссылки на них.
2. Выгрузка из библиотеки набора метаданных, относящихся к статьям различных направлений.
3. Очистка метаданных от некорректных и неполных образцов.
4. Выгрузка набора статей в формате PDF, соответствующих набору метаданных.
5. Очистка набора статей от невалидных и пустых документов.
6. Извлечение метаданных одинаковых типов с помощью отобранных программных решений.
7. Постобработка результатов извлечения и приведение их к единому формату.
8. Сравнение результатов извлечения с выгруженным ранее набором метаданных, вычисление значений метрик качества извлечения.
9. Анализ полученных значений метрик, сравнение программных решений между собой на основании результатов эксперимента.

При сравнении качества извлечения должны быть использованы заранее извлеченные метаданные, позволяющие определить, какой результат работы программных решений может считаться правильным. На основании полученных значений метрик для всех типов метаданных должны быть рассчитаны точность, полнота и F_1 -мера, которые и послужат итоговой оценкой качества извлечения и основанием для выводов о сравнительных возможностях исследуемых программных решений.

3. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА ПРОГРАММНЫХ РЕШЕНИЙ

На первом этапе эксперимента была осуществлена загрузка документов статей и соответствующих им метаданных. В качестве источника метаданных был

выбран Semantic Scholar. Этот интернет-ресурс предоставляет широкие возможности для загрузки метаданных научных статей по различным направлениям и гибкие возможности настройки выгружаемых метаданных. Среди прочего ресурс предоставляет ссылку на исходный документ статьи в формате PDF. При настройке загрузки есть возможность выбрать 23 различных предметных области для статей. При первоначальной загрузке метаданных с помощью предлагаемого этим ресурсом API было загружено 230 000 наборов метаданных, по 10 000 для каждой из предложенных областей, затем из этих наборов были исключены те, в которых присутствовали пустые значения для рассматриваемых типов метаданных. Далее для оставшихся наборов с помощью содержащихся в них ссылок была осуществлена загрузка полных текстов статей в формате PDF. После загрузки из выборки были исключены поврежденные или пустые PDF-файлы. Итоговая выборка наборов метаданных с соответствующим ими PDF-документами составила 112 457 записей, т. е. около 5000 записей на каждую из рассматриваемых предметных областей.

И наконец, с помощью CERMINE, GROBID и ScientificPdfParser из текстов статей были извлечены метаданные. Для GROBID использовался интерфейс SciPDF_Parser, совместимый с Python и предоставляющий настройки для извлекаемых полей. Для всех статей извлекались следующие метаданные: название статьи, фамилии авторов, год публикации и аннотация. Такой набор, с одной стороны, опирается на возможности рассматриваемых решений, с другой – позволяет делать выводы об извлечении отличающихся данных: текст из нескольких предложений в аннотации, набор из имен авторов, численные значения даты. Для GROBID и ScientificPdfParser данные были извлечены в формате JSON, для CERMINE – в формате sermxml, а затем также переведены в формат JSON. Это было обусловлено тем, что наборы метаданных, полученных от Semantic Scholar, также представлены в формате JSON.

Для оценки правильности извлечения метаданных использовались две метрики сходства. Для длинных строк, таких как название статьи и аннотация, использовалась метрика сходства на основе расстояния Левенштейна [10]. Для имен авторов и даты публикации использовалась метрика сходства на основе расстояния Джаро–Винклера, как более подходящая для коротких строк с учетом сравнения совпадающего префикса, в том числе для сравнения имен [11].

Метрика сходства на основе расстояния Левенштейна для строк s_1 и s_2 определяется по формуле

$$\text{sim}_{\text{Lev}}(s_1, s_2) = 1 - \frac{\text{Lev}(s_1, s_2)}{\max(|s_1|, |s_2|)}$$

где $\text{Lev}(s_1, s_2)$ – расстояние Левенштейна для строк s_1 и s_2 , а $|s_1|$ и $|s_2|$ – длины строк.

Метрика сходства на основе расстояния Джаро–Винклера для двух строк s_1 и s_2 определяется по формуле

$$\text{sim}_{\text{JW}}(s_1, s_2) = 1 - (d_j(s_1, s_2) + (l \cdot p \cdot (1 - d_j(s_1, s_2)))),$$

где $d_j(s_1, s_2)$ – расстояние Джаро для строк s_1 и s_2 , l – длина совпадающего префикса (максимум до 4 символов), p – коэффициент масштабирования. Стандартным значением для p является 0.1, оно и было взято для вычислений в рамках этого исследования.

Для определения качества извлечения метаданных необходимо было определить пороговое значение для оценок, получаемых с помощью метрик сходства, после которого извлечение считалось бы корректным. Это обусловлено тем, что существует некоторое отклонение от точного совпадения, обусловленное незначительными неточностями (отсутствием пробелов, нестандартными текстовыми символами, знаками препинания), в рамках которого данные все еще можно считать корректными для дальнейшего использования. Правильное выявление этого порога является важной задачей, поскольку именно на нем строится итоговая оценка качества извлечения метаданных и, как следствие, работы всего программного решения. Алгоритм выявления порогового значения для обеих метрик сходств был одинаков и состоял из следующих шагов.

1. Для поиска порогового значения t берется интервал $[0.5, 1]$, где 1 означает полное совпадение. Интервал разбивается с шагом 0.05, таким образом рассматривается $t \in \{0.05x \mid x \in \{10, 11, \dots, 20\}\}$.
2. Для каждого возможного порогового значения t из общего числа статей выбираются те, у которых вычисленное значение метрики сходства попадает в интервал $[t; t + 0.01]$. Отбор производится таким образом, чтобы в вы-

борке были равным образом представлены статьи из всех предметных областей, извлеченные всеми рассматриваемыми программными решениями.

3. Извлеченные метаданные, соответствующие каждому набору статей, рассматриваются на предмет количества статей с некорректно извлеченными метаданными v_e и его соотношения с общим количеством статей v_a .

4. Итоговым пороговым значением t выбирается то, при котором показатель

$$v = \frac{v_a - v_e}{v_a} \text{ начинал превышать } 0.95.$$

Для выявления порога t_{Lev} по метрике сходства на основе расстояния Левенштейна использовалось название статьи, для порога t_{JW} по метрике сходства на основе расстояния Джаро-Винклера – дата публикации.

На рис. 1 приведен график, показывающий зависимость v от пороговых значений t_{Lev} и t_{JW} .

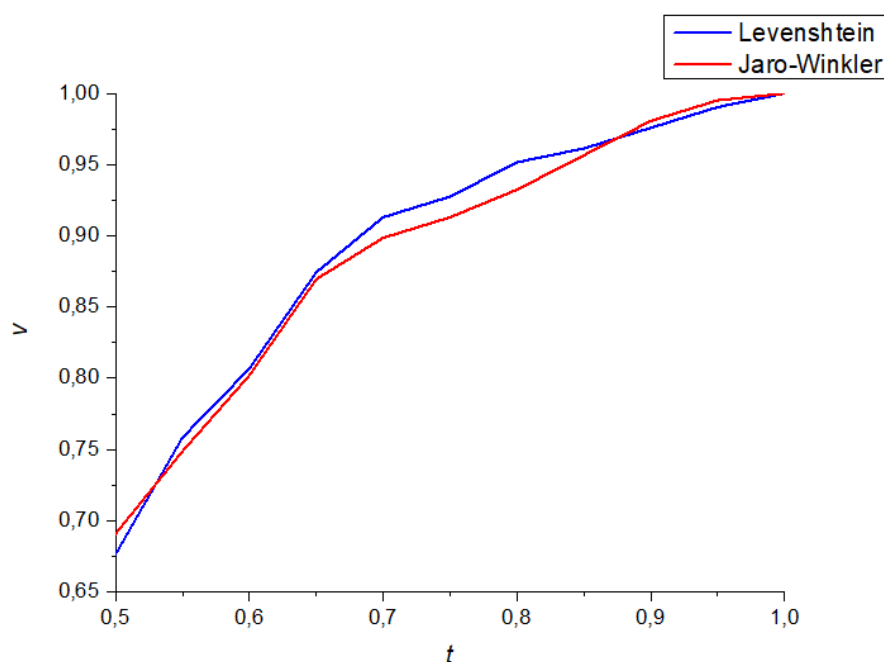


Рис. 1. Зависимость v от пороговых значений t_{Lev} для метрики сходства на основе расстояния Левенштейна и t_{JW} для метрики сходства на основе расстояния Джаро-Винклера

Таким образом, по результатам определения порога были определены $t_{Lev} = 0.8$ и $t_{JW} = 0.85$. Метаданные, для которых значение указанной метрики

сходства превышает пороговое значение, считались корректно извлеченными при вычислении значений метрик качества извлечения.

В качестве метрик качества извлечения использовались полнота (recall), точность (precision) и F₁-мера. Для вычисления точности и полноты было использовано разбиение результатов с помощью матрицы спутанности, согласно которой результаты определения принадлежности к классу (в рассматриваемом случае результаты извлечения метаданных) делятся на положительные TP, ложноположительные FP, отрицательные TN и ложноотрицательные FN. На основе этого полноту и точность можно выразить так:

$$\text{recall} = \frac{TP}{TP+FN}, \text{ precision} = \frac{TP}{TP+FP}.$$

Следует пояснить, что в рассмотренном случае представляют собой результаты по матрице ошибок. TP – это количество статей с корректно извлеченными метаданными, то есть такими, для которых значение метрики сходства превысило вычисленный порог t_{Lev} или t_{JW} соответственно. TP + FP – это количество всех статей, для которых результатом извлечения является непустой ответ, независимо от корректности. TP + FN – это количество статей, у которых образцовые данные являются непустыми, то есть общее количество исследуемых статей. F1-мера представляет собой среднее гармоническое для точности и полноты и определяется по формуле

$$F_1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}.$$

Определение значений по рассмотренным метрикам качества извлечения проводилось по отдельности для каждой представленной предметной области и каждого выбранного типа метаданных.

4. РЕЗУЛЬТАТЫ

На рис. 2–5 представлены рассчитанные на основании результатов эксперимента значения F₁-меры для четырех рассматриваемых типов метаданных в применении к 23 предметным областям для трех исследуемых решений.

Для названия статьи (рис. 2) CERMINE и GROBID показывают схожие результаты, без явного превосходства одной системы над другой, в то время как ScientificPdfParser несколько отстает от этих систем.

Для аннотаций (рис. 3) GROBID показывает существенно лучший результат, чем две другие системы. При этом GROBID показывает худший результат для даты публикации статьи (рис. 4).

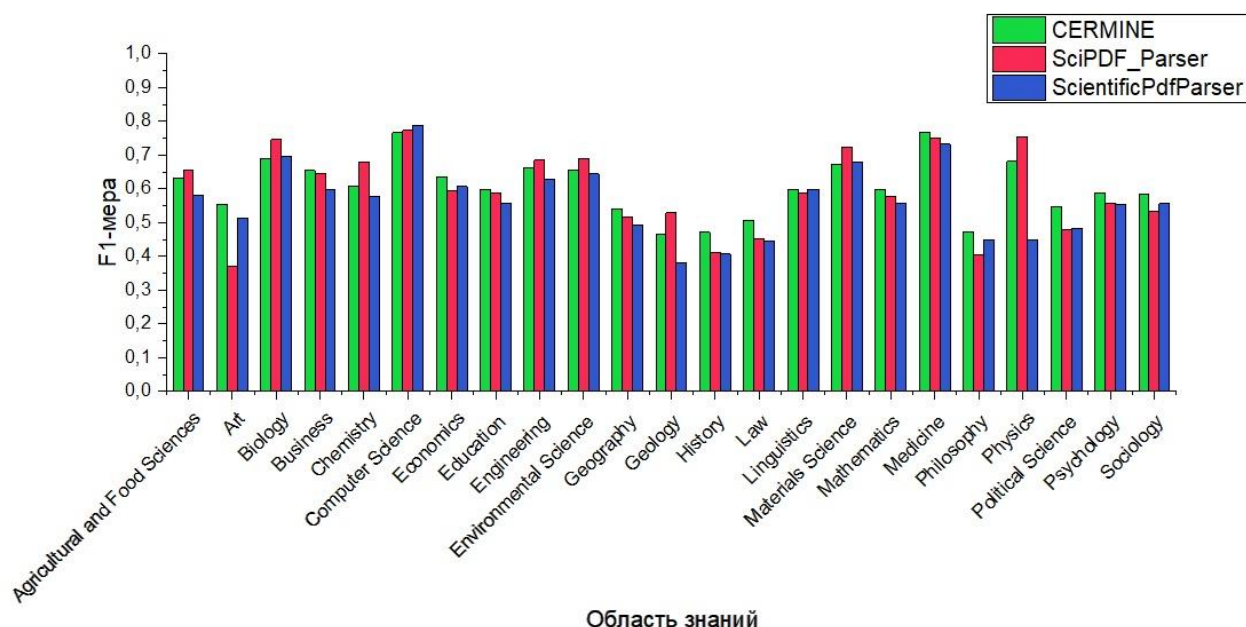


Рис. 2. Значения F_1 -меры для извлечения названия статьи

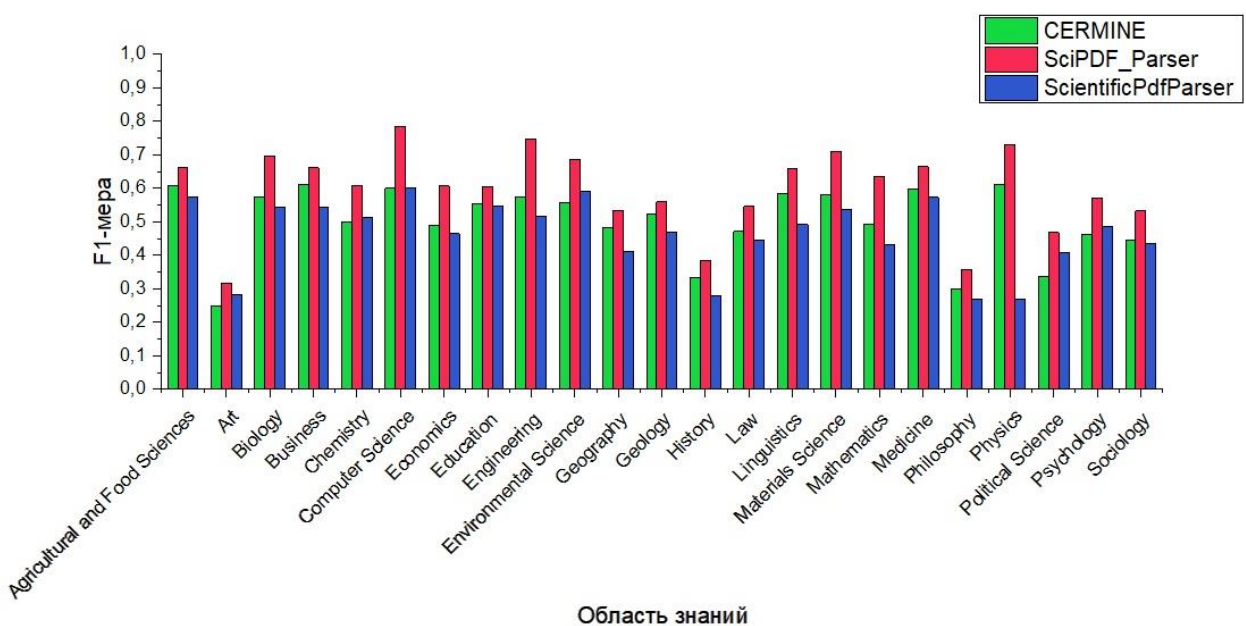


Рис. 3. Значения F_1 -меры для извлечения аннотации статьи

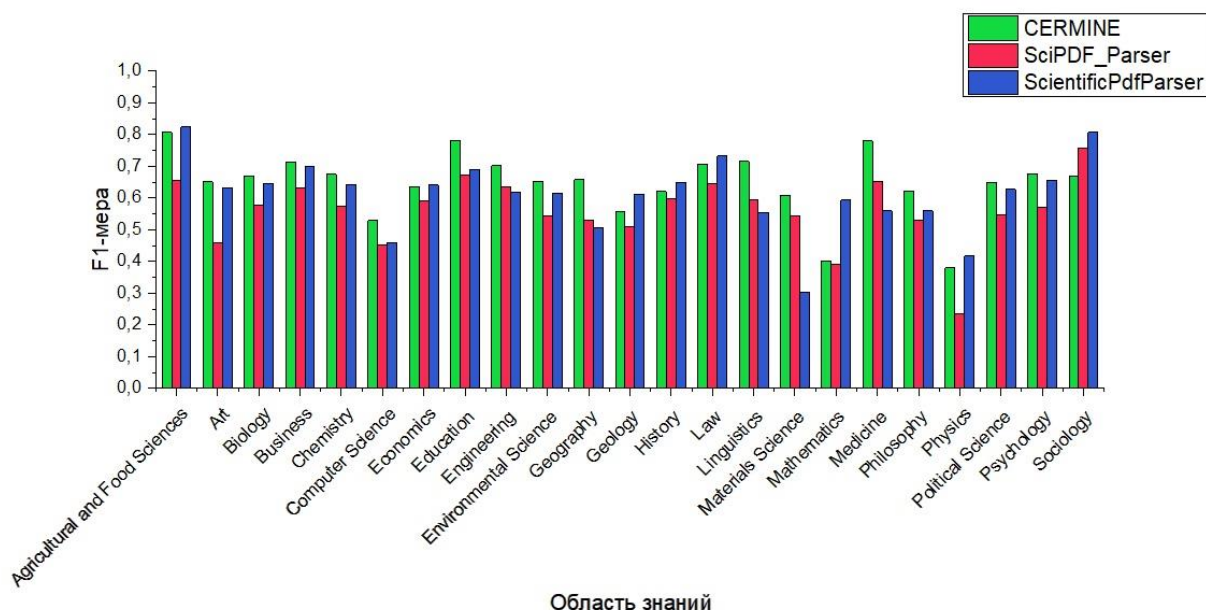


Рис. 4. Значения F_1 -меры для извлечения даты публикации статьи

При извлечении имен авторов GROBID также показывает самый низкий результат, однако в этом случае результаты всех трех систем оказались относительно невысокими (рис. 5).

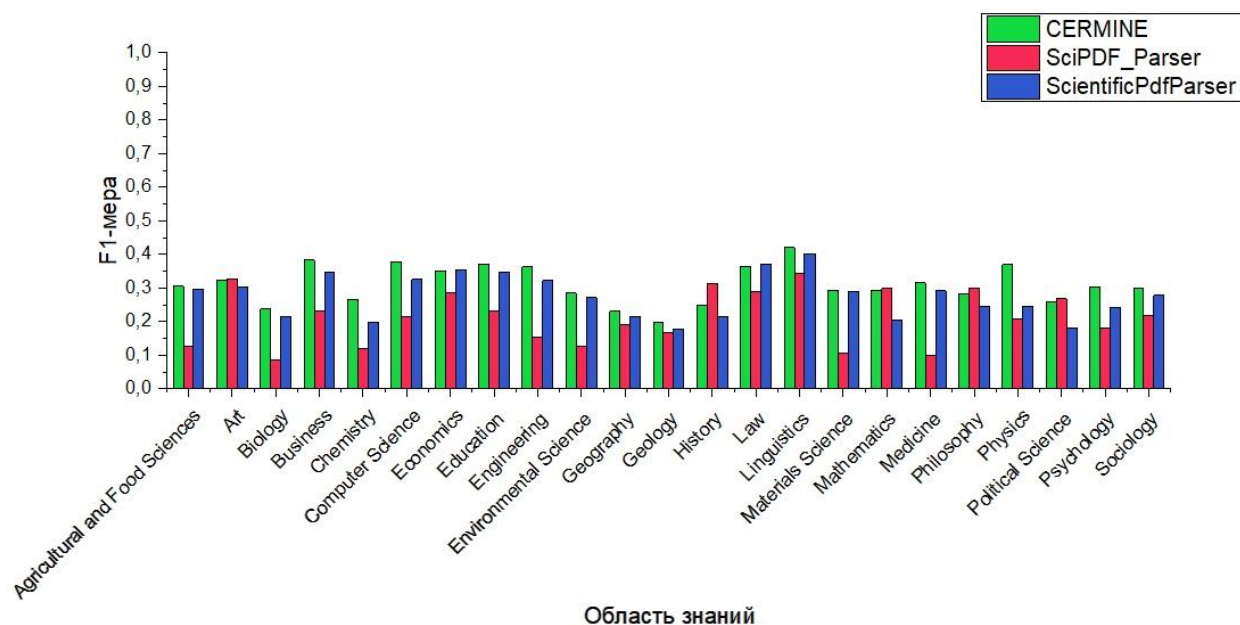


Рис. 5. Значения F_1 -меры для извлечения имен авторов статьи

Одновременно с извлечением имен авторов была произведена оценка корректности их распознавания. Рассмотрение этого фактора обусловлено тем, что

в ряде случаев системам удастся выделить не всех авторов из указанных в статье, или же, наоборот, система дает ряд ложноположительных распознаваний на основании неверного разбиения, определив как отдельных авторов инициалы или же указанные звания и аффилиации авторов. Результаты этой оценки, где метрики основаны на точном совпадении количества авторов, приведены на рис. 6.

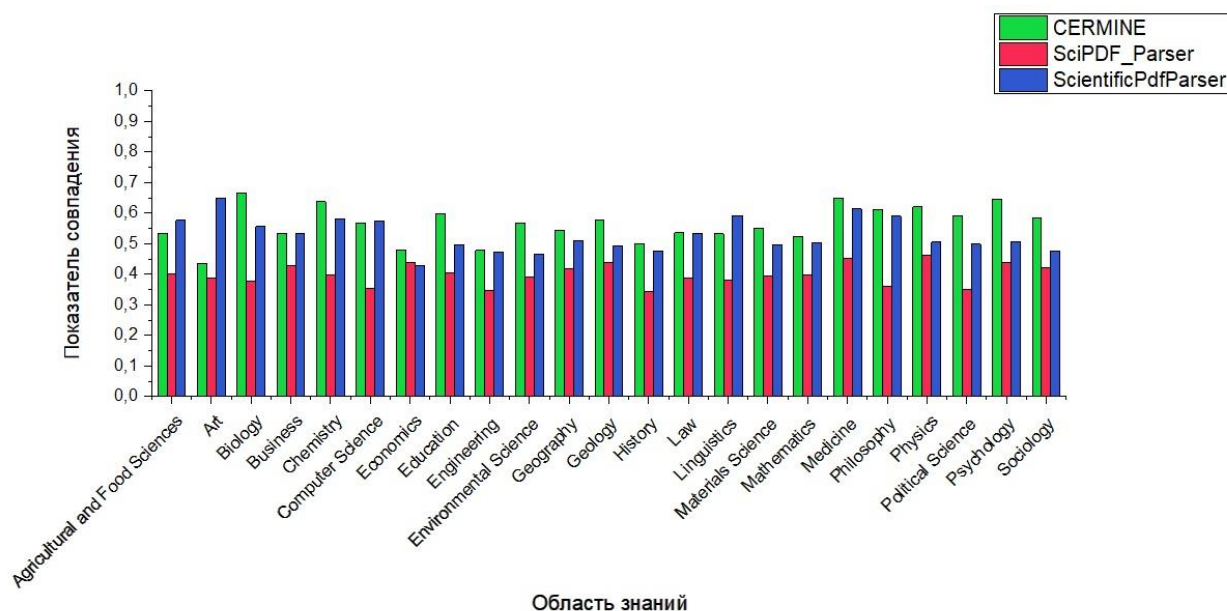


Рис. 6. Значения F_1 -меры для корректности распознавания количества авторов

Все три решения показали относительно невысокий результат для правильного извлечения количества авторов. Это позволяет предполагать, что низкие показатели качества извлечения имен авторов обусловлены не только ошибками в извлечении настоящих имен как строк, но и большим количеством ошибок в распознавании элементов текста как настоящих имен (ложные распознавания и т. д.). Это говорит о необходимости дополнительной настройки систем в области извлечения имен авторов или же дополнительной постобработки извлеченных имен с целью удаления элементов, некорректно включенных в имена. Следует отметить, что эта связь является предполагаемой, поскольку в рамках настоящего исследования не проводились исследования, которые экспериментально выявили бы степень влияния некорректно выделенных и разделенных имен на общее число ошибок при извлечении имен авторов.

Средние значения F_1 -меры для различных метаданных для всех предметных областей были рассчитаны и приведены в табл. 2.

Таблица 2. Среднее значение F_1 -меры для всех рассмотренных предметных областей

Тип метаданных	CERMINE	GROBID	ScientificPdfParser
Название	0.6061	0.5954	0.5641
Аннотация	0.5011	0.5965	0.4644
Дата публикации	0.6461	0.5603	0.6278
Имена авторов	0.3102	0.212	0.2754

5. ВЫБОР ПРОГРАММНОГО РЕШЕНИЯ ИЗВЛЕЧЕНИЯ МЕТАДАНЫХ НА ОСНОВЕ ЭКСПЕРИМЕНТАЛЬНОЙ ОЦЕНКИ

Для определенных областей полученные оценки качества всех трех решений в среднем могут быть ниже, чем для других областей. Так, программные решения показывают наиболее низкие результаты для извлечения названия и аннотации в областях Art, History и Philosophy. Для Biology, Geology и Geography они показывают относительно низкие результаты для всех типов метаданных, в то время как для Economic, Education, Engineering, Law и Linguistics показывают стабильно более высокие результаты. Для Physics программные решения показывают высокие результаты для всех типов метаданных, кроме даты публикации статьи. Это позволяет говорить о том, что различие в результатах может быть не случайным, а обусловлено различными структурой и оформлением статей в различных предметных областях. В связи с этим при работе с конкретной предметной областью целесообразно ориентироваться на качество работы программного решения для нее, а не на средние показатели качества. Выбор программного решения, показывающего наиболее высокое качество извлечения, играет важную роль для решения любой конкретной задачи, одним из этапов которой является извлечение метаданных.

Для выбора программного решения в конкретных условиях предлагается использовать интегральную оценку качества на основе взвешенной суммы. Пусть дано n наборов статей $A_i \subseteq X$, $i \in [1, n]$, полученных для соответствующих предметных областей, где X – множество всех извлеченных статей, и m типов метаданных $d_j \in D$, $j \in [1, m]$, где D – множество типов извлекаемых метаданных.

Для каждого типа метаданных зададим метрику сходства и порог $t_j \in \{t_{\text{Lev}}, t_{\text{JW}}\}, j \in [1, m]$, превышение которого будет означать корректное распознавание. Пусть также даны веса w_i , отражающие значимость предметных областей, и веса u_j , отражающие значимость извлекаемых метаданных, такие, что $\sum_{i=1}^n w_i = 1$ и $\sum_{j=1}^m u_j = 1$.

Пусть дано o программных решений $s_k \in S, k \in [1, o]$, где S – множество рассматриваемых программных решений. Тогда эффективным будет программное решение $s^* \in S$, которое позволит максимизировать интегральную оценку качества:

$$s^* = \arg \max_{s_k \in S} \left(\sum_{j=1}^m \left(u_j \sum_{i=1}^n w_i f(s_k, A_i, d_j, \text{sim}_j, t_j, g) \right) \right),$$

где $f(s_k, A_i, d_j, \text{sim}_j, t_j, g)$ – функция, используемая для вычисления значения метрики качества извлечения g (в данном случае, F_1 -меры) для конкретных программного решения s_k , набора статей A_i и типа метаданных d_j при используемой метрике сходства sim_j и заданном для нее пороговом значении t_j .

Рассмотрим пример определения эффективного программного решения с использованием интегральных оценок. Дано $n = 23$ набора статей $A_i, i \in [1, n]$, соответствующих рассмотренным ранее предметным областям, $m = 4$ рассматриваемых типа метаданных $d_j, j \in [1, m]$ (название d_1 , аннотация d_2 , дата публикации d_3 , имена авторов d_4) и $o = 3$ программных решения $s_k, k \in [1, o]$ для оценки (CERMINE, GROBID, ScientificPdfParser).

Для каждого типа метаданных установим метрики сходства и вычисленные ранее пороговые значения. Для названия статьи и аннотации $\text{sim}_1 = \text{sim}_2 = \text{sim}_{\text{Lev}}, t_1 = t_2 = t_{\text{Lev}} = 0.8$; для даты публикации и имен авторов $\text{sim}_3 = \text{sim}_4 = \text{sim}_{\text{JW}}, t_3 = t_4 = t_{\text{JW}} = 0.85$.

Пусть веса для предметных областей распределяются следующим образом: Engineering, Mathematics, Physics $w_1 = w_2 = w_3 = 0.1$; Materials Science, Geology,

Chemistry, Medicine $w_4 = w_5 = w_6 = w_7 = 0.07$; Computer Science, Environmental Science $w_8 = w_9 = 0.05$; Biology $w_{10} = 0.04$; Agricultural and Food Science, Economics, Geography, Business, Education $w_{11} = w_{12} = w_{13} = w_{14} = w_{15} = 0.03$; Sociology, Psychology, Political Science, Law, Linguistics $w_{16} = w_{17} = w_{18} = w_{19} = w_{20} = 0.02$; Art, History, Philosophy $w_{21} = w_{22} = w_{23} = 0.01$. С использованием этих весов получим промежуточные оценки для четырех рассматриваемых типов метаданных (табл. 3) и программных решений.

Таблица 3. Промежуточные оценки качества программных решений для четырех типов метаданных

Тип метаданных	CERMINE	GROBID	ScientificPdfParser
Название	0.6318	0.6462	0.5801
Аннотация	0.5396	0.6486	0.4793
Дата публикации	0.6127	0.5286	0.5772
Имена авторов	0.3106	0.1879	0.2665

Пусть для типов метаданных веса распределяются следующим образом: имена авторов $u_1 = 0.4$; название $u_2 = 0.3$; дата публикации $u_3 = 0.2$; аннотация $u_4 = 0.1$. Тогда для итоговой оценки качества работы программных решений получим следующие значения: CERMINE – 0.4912; GROBID – 0.4396; ScientificPdfParser – 0.444, и соответственно, рейтинг программ выстраивается следующим образом:

CERMINE $\rightarrow$ ScientificPdfParser $\rightarrow$ GROBID.

CERMINE по интегральной оценке качества показывает эффективность выше, чем GROBID, на 10.5%, и выше, чем ScientificPdfParser, на 9.6%. Таким образом, получен рейтинг программных решений с учетом выбранных весов для предметных областей и типов метаданных для представленного примера.

ЗАКЛЮЧЕНИЕ

Представлена методика сравнения программных решений распознавания текстов научных публикаций в формате PDF по качеству извлечения метаданных. Для извлечения рассмотрены четыре типа метаданных: название, аннотация,

дата публикации, имена авторов. Проанализированы три программных решения для извлечения метаданных: CERMINE, GROBID и ScientificPdfParser.

Проведен эксперимент по сравнению решений на основе подготовленного набора данных и выбора программных решений в условиях заданных приоритетов для предметных областей и типов метаданных. В ходе эксперимента были установлены и сконфигурированы программные решения, примененные для извлечения метаданных. Получена оценка качества извлечения на основании сравнения результатов с выгрузкой метаданных из сервиса Semantic Scholar. При извлечении общий набор статей был разделен на 23 предметных области, что позволило в дальнейшем сравнить качество извлечения для этих областей и сделать предположение о наличии связи качества извлечения от специфики оформления статей, присущих той или иной предметной области.

При выборе программных решений предложено использовать интегральную оценку на основе взвешенной суммы. Рассмотрен пример присвоения приоритетов предметным областям и типам метаданных, получен рейтинг программных решений в рамках заданных условий.

Разработанная методика имеет прикладное значение. Предложенную методику и полученные данные о качестве извлечения для различных программных средств можно использовать для выбора программного решения в условиях решения конкретной задачи.

Дальнейшие исследования могут быть посвящены особенностям функционирования программных решений, показателям их временной эффективности и вычислительной ресурсоемкости. Можно рассмотреть также построение ансамблей из программных решений извлечения метаданных с целью повышения качества извлечения.

СПИСОК ЛИТЕРАТУРЫ

1. *Qayyum F., Afzal M.T.* Identification of important citations by exploiting research articles' metadata and cue-terms from content // *Scientometrics*. 2019. Vol. 118. P. 21–43.
2. *Liu X., Zhang J., Guo C.* Full-text citation analysis: A new method to enhance scholarly networks // *Journal of the American Society for Information Science and Technology*. 2013. Vol. 64. No. 9. P. 1852–1863.

3. *Saier T., Färber M.* unarXive: a large scholarly data set with publications' full-text, annotated in-text citations, and links to metadata // *Scientometrics*. 2020. Vol. 125. No. 3. P. 3085–3108.

4. *Safder I. et al.* Deep learning-based extraction of algorithmic metadata in full-text scholarly documents // *Information processing and management*. 2020. Vol. 57. No. 6. Article no. 102269.

5. *O'Leary N.A. et al.* Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets // *Scientific data*. 2024. Vol. 11. No. 1. Article no. 732.

6. *Safder I., Hassan S.U.* Bibliometric-enhanced information retrieval: a novel deep feature engineering approach for algorithm searching from full-text publications // *Scientometrics*. 2019. Vol. 119. P. 257–277.

7. *Joshi B., Symeonidou A., Danish S.M., Hermesen F.* An End-to-End Pipeline for Bibliography Extraction from Scientific Articles // *Proceedings of the Second Workshop on Information Extraction from Scientific Publications*. 2023. P. 101–106.

8. *Ma A. et al.* A deep-learning based citation count prediction model with paper metadata semantic features // *Scientometrics*. 2021. Vol. 126. No. 8. P. 6803–6823.

9. *Lo K. et al.* PaperMage: A Unified Toolkit for Processing, Representing, and Manipulating Visually-Rich Scientific Documents // *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2023. P. 495–507.

10. *Po D.K.* Similarity based information retrieval using Levenshtein distance algorithm // *International Journal of Advances in Scientific Research and Engineering*. 2020. Vol. 6. No. 04. P. 6–10.

11. *Nurcahyawati V., Mustaffa Z.* Online Media as a Price Monitor: Text Analysis using Text Extraction Technique and Jaro-Winkler Similarity Algorithm // *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*. IEEE, 2020. P. 1–6.

12. *Foppiano L. et al.* Automatic extraction of materials and properties from superconductors scientific literature // *Science and Technology of Advanced Materials: Methods*. 2023. Vol. 3. No. 1. Article no. 2153633.

13. *Petersen T. et al.* Geo-quantities: A framework for automatic extraction of measurements and spatial context from scientific documents // Proceedings of the 17th International Symposium on Spatial and Temporal Databases. 2021. P. 166–169.

14. *Chraibi A. et al.* Extraction of measurements from medical reports // 10ème conférence Francophone en Gestion et Ingénierie des Systèmes Hospitaliers, GISEH2020. 2020.

15. *Haviana S.F.C., Subroto I. M.I.* Obtaining Reference's Topic Congruity in Indonesian Publications using Machine Learning Approach // 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). IEEE. 2019. P. 428–431.

16. *Ermakova L. Bordignon F., Turenne N., Noel M.* Is the Abstract a Mere Teaser? Evaluating generosity of article abstracts in the environmental sciences // Frontiers in Research Metrics and Analytics. 2018. Vol. 3. Article no. 16.

17. *El-Ebshihy A. et al.* A platform for argumentative zoning annotation and scientific summarization // Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022. P. 4843–4847.

18. *Choi W. et al.* Building an annotated corpus for automatic metadata extraction from multilingual journal article references // PloS one. 2023. Vol. 18. No. 1. Article no. E0280637.

19. *Krause J. et al.* Bootstrapping multilingual metadata extraction: a showcase in cyrillic // Proceedings of the Second Workshop on Scholarly Document Processing. 2021. P. 66–72.

20. *Shapiro I., Saier T., Färber M.* Sequence Labeling for Citation Field Extraction from Cyrillic Script References // Proceedings of the Workshop on Scientific Document Understanding; co-located with 36th AAAI Conference on Artificial Intelligence (AAAI 2022). 2022.

21. *Indrawati A., Yoganingrum A., Yuwono P.* Evaluating the quality of the indonesian scientific journal references using ParsCit, CERMINE and GROBID // Library Philosophy and Practice. 2019. P. 1–14.

22. *Meuschke N. et al.* A benchmark of pdf information extraction tools using a multi-task and multi-domain evaluation framework for academic documents // International Conference on Information. Cham: Springer Nature Switzerland, 2023. P. 383–405.

23. *Guo Z., Jin H.* Reference metadata extraction from scientific papers // 2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies. IEEE. 2011. P. 45–49.

24. *Beel J., Langer S., Genzmehr M., M"uller C.* Docear's PDF inspector: title extraction from PDF files // Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries. New York, NY, USA: ACM, 2013. P. 443–444.

25. *Jensen Z. et al.* A machine learning approach to zeolite synthesis enabled by automatic literature data extraction // ACS central science. 2019. Vol. 5. No. 5. P. 892–899.

26. *Färber M., Albers A., Schüber F.* Identifying used methods and datasets in scientific publications // Proceedings of the Workshop on Scientific Document Understanding co-located with 35th AAAI Conference on Artificial Inteligence (AAAI 2021). 2021.

27. *Suryawati E., Widyantoro D.H.* Combination of heuristic, rule-based and machine learning for bibliography extraction // 2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME). IEEE. 2017. P. 276–281.

28. *Tkaczyk D. et al.* CERMINE: automatic extraction of structured metadata from scientific literature // International Journal on Document Analysis and Recognition (IJДАР). 2015. Vol. 18. P. 317–335.

29. *Romary L., Lopez P.* Grobid-information extraction from scientific publications // ERCIM News. 2015. Vol. 100.

30. *Councill I.G., Giles C.L., Kan M.Y.* ParsCit: an Open-source CRF Reference String Parsing Package // Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008. 2008. Vol. 8. P. 661–667.

31. *Prasad A., Kaur M., Kan M.Y.* Neural ParsCit: a deep learning-based reference string parser // International journal on digital libraries. 2018. Vol. 19. P. 323–337.

32. *Constantin A., Pettifer S., Voronkov A.* PDFX: fully-automated PDF-to-XML conversion of scientific literature // Proceedings of the 2013 ACM symposium on Document engineering. 2013. P. 177–180.

PROCEDURE FOR COMPARING TEXT RECOGNITION SOFTWARE SOLUTIONS FOR SCIENTIFIC PUBLICATIONS BY THE QUALITY OF METADATA EXTRACTION

I. I. Kuznetsov¹ [0009-0001-6287-8295], O. P. Novikov² [0009-0009-3494-3799],
D. Y. Ilin³ [0000-0002-0241-2733]

^{1, 2}A.N. Kosygin Moscow State Textile University, Moscow, 115035 Russia;

³MIREA – Russian Technological University, Moscow, 119454 Russia

¹iliya-kuznetsov@mail.ru, ²novikovop55@rambler.ru, ³i@dmitryilin.com

Abstract

Metadata of scientific publications are used to build catalogs, determine the citation of publications, and perform other tasks. Automation of metadata extraction from PDF files provides means to speed up the execution of the designated tasks, while the possibility of further use of the obtained data depends on the quality of extraction. Existing software solutions were analyzed, after which three of them were selected: GROBID, CERMINE, ScientificPdfParser. A procedure for comparing software solutions for recognizing texts of scientific publications by the quality of metadata extraction is proposed. Based on the procedure, an experiment was conducted to extract 4 types of metadata (title, abstract, publication date, author names). To compare software solutions, a dataset of 112,457 publications divided into 23 subject areas formed on the basis of Semantic Scholar data was used. An example of choosing an effective software solution for metadata extraction under the conditions of specified priorities for subject areas and types of metadata using a weighted sum is given. It was determined that for the given example CERMINE shows efficiency 10.5% higher than GROBID and 9.6% higher than ScientificPdfParser.

Keywords: *text recognition, scientific publications, metadata, data extraction quality, procedure.*

REFERENCES

1. Qayyum F., Afzal M.T. Identification of important citations by exploiting re-search articles' metadata and cue-terms from content // *Scientometrics*. 2019. Vol. 118. P. 21–43.
2. Liu X., Zhang J., Guo C. Full-text citation analysis: A new method to enhance scholarly networks // *Journal of the American Society for Information Science and Technology*. 2013. Vol. 64. No. 9. P. 1852–1863.
3. Saier T., Färber M. unarXive: a large scholarly data set with publications' full-text, annotated in-text citations, and links to metadata // *Scientometrics*. 2020. Vol. 125. No. 3. P. 3085–3108.
4. Safder I. et al. Deep learning-based extraction of algorithmic metadata in full-text scholarly documents // *Information processing and management*. 2020. Vol. 57. No. 6. Article no. 102269.
5. O'Leary N.A. et al. Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets // *Scientific data*. 2024. Vol. 11. No. 1. Article no. 732.
6. Safder I., Hassan S.U. Bibliometric-enhanced information retrieval: a novel deep feature engineering approach for algorithm searching from full-text publications // *Scientometrics*. 2019. Vol. 119. P. 257–277.
7. Joshi B., Symeonidou A., Danish S.M., Hermesen F. An End-to-End Pipeline for Bibliography Extraction from Scientific Articles // *Proceedings of the Second Workshop on Information Extraction from Scientific Publications*. 2023. P. 101–106.
8. Ma A. et al. A deep-learning based citation count prediction model with paper metadata semantic features // *Scientometrics*. 2021. Vol. 126. No. 8. P. 6803–6823.
9. Lo K. et al. PaperMage: A Unified Toolkit for Processing, Representing, and Manipulating Visually-Rich Scientific Documents // *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. 2023. P. 495–507.
10. Po D.K. Similarity based information retrieval using Levenshtein distance algorithm // *International Journal of Advances in Scientific Research and Engineering*. 2020. Vol. 6. No. 04. P. 6–10.

11. *Nurcahyawati V., Mustaffa Z.* Online Media as a Price Monitor: Text Analysis using Text Extraction Technique and Jaro-Winkler Similarity Algorithm // 2020 Emerging Technology in Computing, Communication and Electronics (ETCCE). IEEE, 2020. P. 1–6.

12. *Foppiano L. et al.* Automatic extraction of materials and properties from superconductors scientific literature // Science and Technology of Advanced Materials: Methods. 2023. Vol. 3. No. 1. Article no. 2153633.

13. *Petersen T. et al.* Geo-quantities: A framework for automatic extraction of measurements and spatial context from scientific documents // Proceedings of the 17th International Symposium on Spatial and Temporal Databases. 2021. P. 166–169.

14. *Chraibi A. et al.* Extraction of measurements from medical reports // 10ème conférence Francophone en Gestion et Ingénierie des Systèmes Hospitaliers, GISEH2020. 2020.

15. *Haviana S.F.C., Subroto I. M.I.* Obtaining Reference's Topic Congruity in Indonesian Publications using Machine Learning Approach // 2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). IEEE. 2019. P. 428–431.

16. *Ermakova L. Bordignon F., Turenne N., Noel M.* Is the Abstract a Mere Teaser? Evaluating generosity of article abstracts in the environmental sciences // Frontiers in Research Metrics and Analytics. 2018. Vol. 3. Article no. 16.

17. *El-Ebshihy A. et al.* A platform for argumentative zoning annotation and scientific summarization // Proceedings of the 31st ACM International Conference on Information & Knowledge Management. 2022. P. 4843–4847.

18. *Choi W. et al.* Building an annotated corpus for automatic metadata extraction from multilingual journal article references // PloS one. 2023. Vol. 18. No. 1. Article no. E0280637.

19. *Krause J. et al.* Bootstrapping multilingual metadata extraction: a showcase in cyrillic // Proceedings of the Second Workshop on Scholarly Document Processing. 2021. P. 66–72.

20. *Shapiro I., Saier T., Färber M.* Sequence Labeling for Citation Field Extraction from Cyrillic Script References // Proceedings of the Workshop on Scientific Document Understanding; co-located with 36th AAAI Conference on Artificial Intelligence (AAAI 2022). 2022.

21. *Indrawati A., Yoganingrum A., Yuwono P.* Evaluating the quality of the indonesian scientific journal references using ParsCit, CERMINE and GROBID // *Library Philosophy and Practice*. 2019. P. 1–14.
22. *Meuschke N. et al.* A benchmark of pdf information extraction tools using a multi-task and multi-domain evaluation framework for academic documents // *International Conference on Information*. Cham: Springer Nature Switzerland, 2023. P. 383–405.
23. *Guo Z., Jin H.* Reference metadata extraction from scientific papers // *2011 12th International Conference on Parallel and Distributed Computing, Applications and Technologies*. IEEE. 2011. P. 45–49.
24. *Beel J., Langer S., Genzmehr M., Müller C.* Docear's PDF inspector: title extraction from PDF files // *Proceedings of the 13th ACM/IEEE-CS joint conference on Digital libraries*. New York, NY, USA: ACM, 2013. P. 443–444.
25. *Jensen Z. et al.* A machine learning approach to zeolite synthesis enabled by automatic literature data extraction // *ACS central science*. 2019. Vol. 5. No. 5. P. 892–899.
26. *Färber M., Albers A., Schüber F.* Identifying used methods and datasets in scientific publications // *Proceedings of the Workshop on Scientific Document Understanding co-located with 35th AAAI Conference on Artificial Inteligence (AAAI 2021)*. 2021.
27. *Suryawati E., Widiantoro D.H.* Combination of heuristic, rule-based and machine learning for bibliography extraction // *2017 5th International Conference on Instrumentation, Communications, Information Technology, and Biomedical Engineering (ICICI-BME)*. IEEE. 2017. P. 276–281.
28. *Tkaczyk D. et al.* CERMINE: automatic extraction of structured metadata from scientific literature // *International Journal on Document Analysis and Recognition (IJ DAR)*. 2015. Vol. 18. P. 317–335.
29. *Romary L., Lopez P.* Grobid-information extraction from scientific publications // *ERCIM News*. 2015. Vol. 100.
30. *Councill I.G., Giles C.L., Kan M.Y.* ParsCit: an Open-source CRF Reference String Parsing Package // *Proceedings of the 6th International Conference on Language Resources and Evaluation, LREC 2008*. 2008. Vol. 8. P. 661–667.

31. Prasad A., Kaur M., Kan M.Y. Neural ParsCit: a deep learning-based reference string parser // International journal on digital libraries. 2018. Vol. 19. P. 323–337.

32. Constantin A., Pettifer S., Voronkov A. PDFX: fully-automated PDF-to-XML conversion of scientific literature // Proceedings of the 2013 ACM symposium on Document engineering. 2013. P. 177–180.

СВЕДЕНИЯ ОБ АВТОРАХ



КУЗНЕЦОВ Илья Игоревич – аспирант кафедры искусственного интеллекта, прикладной математики и программирования Института информационных технологий и цифровой трансформации, Российский государственный университет им. А.Н. Косыгина (Технологии. Дизайн. Искусство).

Ilya Igorevich KUZNETSOV – postgraduate student of the Department of Artificial Intelligence, Applied Mathematics and Programming, Institute of Information Technologies and Digital Transformation, A. N. Kosygin Moscow State Textile University.

email: iliya-kuznetsov@mail.ru

ORCID: 0009-0001-6287-8295



НОВИКОВ Олег Пантелеевич – доктор технических наук, профессор, профессор кафедры искусственного интеллекта, прикладной математики и программирования Института информационных технологий и цифровой трансформации, Российский государственный университет им. А.Н. Косыгина (Технологии. Дизайн. Искусство).

Oleg Panteleevich NOVIKOV – Doctor of Engineering, professor, professor of the Department of Artificial Intelligence, Applied Mathematics and Programming of the Institute of Information Technology and Digital Transformation, A. N. Kosygin Moscow State Textile University.

email: novikovop55@rambler.ru

ORCID: 0009-0009-3494-3799



ИЛЬИН Дмитрий Юрьевич – кандидат технических наук, доцент кафедры КБ-14 «Цифровые технологии обработки данных» Института кибербезопасности и цифровых технологий, МИРЭА – Российский технологический университет.

Dmitry Yurievich ILIN – Candidate of Sciences (Engineering), associate professor at the department of Data Processing Digital Technologies, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University.

email: i@dmitryilin.com

ORCID: 0000-0002-0241-2733

Материал поступил в редакцию 5 апреля 2025 года

СИГНАТУРНЫЕ МЕТОДЫ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

К. А. Машченко^[0000-0003-0355-6699]

НИЦ «Курчатовский институт» – НИИСИ, г. Москва, 117218, Россия

kirill.mashchenko@niisi.ru

Аннотация

Сигнатурные методы представляют собой мощный инструмент анализа временных рядов, который преобразует их в форму, удобную для задач машинного обучения. В статье рассмотрены основные понятия сигнатуры пути, ее свойства и геометрический смысл, а также методы вычисления для различных типов временных рядов. Приведены примеры применения сигнатурных методов в различных областях, включая финансы, медицину и образование, продемонстрированы их преимущества перед традиционными подходами. Особое внимание уделено генерации синтетических данных на основе сигнатур, что особенно актуально в условиях ограниченного объема исходных данных. Представлены результаты экспериментальных исследований по генерации и предсказанию траекторий цифрового следа обучения студентов, подтверждающие эффективность сигнатурных методов для применения в задачах машинного обучения по анализу и прогнозированию временных рядов.

Ключевые слова: *сигнатура, сигнатурные методы, временные ряды, генерация данных, анализ траекторий, цифровой след.*

ВВЕДЕНИЕ

Сигнатурные методы являются одним из эффективных инструментов для обработки и выделения признаков из пути – некоторой функции, аргументом которой является время, например, траектории случайного процесса. С помощью этого инструментария можно преобразовать временные ряды в форму, более подходящую для задач машинного обучения, таких как классификация и прогнозирование.

Понятие сигнатуры пути впервые было введено Ченом [1, 2], она представляет собой последовательность итерированных интегралов, которые отражают

различные свойства пути. Известно, что при выполнении некоторых достаточно общих условий, наложенных на рассматриваемый класс путей, существует взаимно однозначное соответствие между сигнатурами и путями. В рамках различных практических направлений, от задач восстановления траекторий до анализа потоков данных различной природы, было установлено, что достижение приемлемого уровня точности восстановления пути по его сигнатуре зачастую возможно при использовании лишь начальных членов сигнатуры, а именно первых нескольких повторных интегралов. Таким образом, сигнатура временного ряда, как способ кодирования, оказывается инструментом высокоэффективной компрессии информации, заключенной в траектории.

Временные ряды, являющиеся представлением реальных данных, по сути представляются своими значениями в дискретном множестве моментов времени. Таким образом, все рассматриваемые траектории являются кусочно-линейными, что упрощает определение сигнатуры и использование ее свойств.

В работах Чена, указанных выше, рассматривались достаточно регулярные траектории, то есть траектории с ограниченной вариацией, допускающие применение интегралов Римана–Стилтьеса. Для грубых траекторий (например, с показателем Гельдера меньше $1/2$), которые требуют применения стохастических интегралов или интегралов по грубым траекториям, сигнатуры исследовались Лайонсом, Хайрером и др. [3–5]. Оказалось, что ряд результатов, например взаимно однозначное соответствие между траекториями и сигнатурами, остается верным и для грубых траекторий [6, 7].

На протяжении последнего десятилетия наблюдается стремительное проникновение сигнатурных методов в арсенал прикладного машинного обучения. Так, будучи скомбинированными с архитектурами сверточных нейронных сетей, они обеспечили получение первенства в престижном онлайн-соревновании ICDAR 2013 по распознаванию изолированных китайских иероглифов [8], а в сопряджении с моделью градиентного бустинга заняли лидирующую позицию в состязании PhysioNet 2019, сосредоточенном на вычислительных аспектах кардиологической диагностики [9]. Кроме того, эти методы нашли широкое применение в области финансовой математики, в частности при решении задач хеджирования производных инструментов [10]. Для применения методов машинного обучения

в условиях недостаточного количества эмпирических данных была предложена методология генерации дополнительных траекторий, статистически согласованных с оригиналом [11].

Сигнатурные методы также успешно применяются в задачах классификации и регрессии. Например, в работе [12] сигнатуры использованы для анализа финансовых временных рядов, где они позволяют выделять ключевые особенности данных. Авторы установили, что даже небольшого количества коэффициентов сигнатуры достаточно для классификации финансовых потоков и прогнозирования рыночных изменений.

Традиционные методы анализа временных рядов, такие как динамические факторные модели (DFM) и авторегрессионные модели (ARIMA), часто требуют строгих предположений о стационарности и линейности данных. В отличие от них, сигнатурные методы не накладывают таких ограничений. В работе [13] показано, что регрессия на сигнатурах превосходит DFM в задаче прогнозирования ВВП, обеспечивая меньшую ошибку и большую устойчивость к нерегулярностям данных.

Одной из проблем сигнатурных методов является экспоненциальный рост числа сигнатурных членов с увеличением уровня усечения. Для решения этой проблемы применяются методы регуляризации, такие как LASSO и Elastic Net [14], а также методы уменьшения размерности, например метод главных компонент (PCA). В работе [15] обсуждается, как выбор уровня усечения и стандартизация элементов сигнатуры влияют на качество моделей.

МЕТОДЫ И МАТЕРИАЛЫ

Сигнатурные методы

С теоретической точки зрения сигнатурный подход следует квалифицировать как непараметрическую, устойчивую к шумам методику извлечения репрезентативных признаков, способных впоследствии служить входными признаками для моделей обучения. Его фундаментальная сила заключается в способности по параметризованному пути, задающему последовательность наблюдаемых со-

стояний, сформировать компактный, но исчерпывающий набор признаков, аккумулирующих как аналитические, так и геометрические характеристики рассматриваемого процесса.

Определение и свойства пути

Путь в R^d – это непрерывное отображение X из некоторого интервала $[a, b]$ в R^d . Чтобы подчеркнуть зависимость от времени, используют обозначение $X_t = X(t): [a, b] \rightarrow R^d$.

В дальнейшем будем предполагать, что рассматриваемые пути являются достаточно «хорошими» отображениями, а именно, они являются кусочно-дифференцируемыми (вообще говоря, справедливость большинства результатов сохранится, если считать, что пути имеют ограниченную вариацию). Будем называть путь гладким, если он бесконечно дифференцируем.

Интегралом от функции $f: R \mapsto R$ по одномерному пути $X: [a, b] \mapsto R$ называется величина

$$\int_a^b f(X_t) dX_t = \int_a^b f(X_t) \dot{X}_t dt,$$

где последний интеграл является обычным (римановым) интегралом непрерывной ограниченной функции. Обозначение «верхняя точка» здесь и далее использовано для дифференцирования по одной переменной: $\dot{X}_t = dX_t/dt$.

Заметим, что $f(X_t)$ тоже является путем на $[a, b]$, поэтому можно естественным образом определить интеграл от пути по пути.

Интегралом от пути $Y: [a, b] \mapsto R$ по пути $X: [a, b] \mapsto R$ называется величина

$$\int_a^b Y_t dX_t = \int_a^b Y_t \dot{X}_t dt.$$

Определение сигнатуры пути

Обозначим координаты пути $X: [a, b] \mapsto R^d$ как $(X_t^1, \dots, X_t^d)$, где каждая координата $X^i: [a, b] \mapsto R$ является путем. Для всех $i \in \{1, \dots, d\}$ и $t \in [a, b]$ определим величину

$$S(X)_{a,t}^i = \int_{a < s < t} dX_s^i = X_t^i - X_a^i,$$

которая является приращением i -й координаты пути до момента времени t . Отметим, что $S(X)_{a,\cdot}^i$ – это тоже путь.

Теперь определим для любой пары $i, j \in \{1, \dots, d\}$ двойной повторный интеграл

$$S(X)_{a,t}^{i,j} = \int_{a < s < t} S(X)_{a,s}^i dX_s^j = \int_{a < r < s < t} dX_r^i dX_s^j = \int_{a < s < t} \int_{a < r < s} dX_r^i dX_s^j.$$

Можно продолжить по индукции: для любых $k \geq 1$ и набора индексов $i_1, \dots, i_k \in \{1, \dots, d\}$ определим

$$S(X)_{a,t}^{i_1, \dots, i_k} = \int_{a < s < t} S(X)_{a,s}^{i_1, \dots, i_{k-1}} dX_s^{i_k} = \int_{a < t_k < t} \dots \int_{a < t_1 < t_2} dX_{t_1}^{i_1} \dots dX_{t_k}^{i_k}.$$

Величина $S(X)_{a,t}^{i_1, \dots, i_k}$ называется k -кратным повторным интегралом от пути X по индексам $i_1, \dots, i_k$.

Сигнатурой пути $X: [a, b] \mapsto R^d$ называется бесконечный набор $S(X)_{a,b}$ всех повторных интегралов от X :

$$S(X)_{a,b} = (1, S(X)_{a,b}^1, \dots, S(X)_{a,b}^d, S(X)_{a,b}^{1,1}, S(X)_{a,b}^{1,2}, \dots),$$

где первый элемент сигнатуры (соответствующий пустому индексу) по определению считается равным 1, а верхние индексы остальных элементов пробегают набор всевозможных мульти-индексов

$$W = \{(i_1, \dots, i_k) \mid k \geq 1, \quad i_1, \dots, i_k \in \{1, \dots, d\}\}.$$

Множество W называется множеством слов в алфавите $A = \{1, \dots, d\}$. Конечный набор чисел $S(X)_{a,b}^{i_1, \dots, i_k}$ для всевозможных мультииндексов длины k будем называть k -м уровнем сигнатуры.

Пример 1. Рассмотрим произвольный одномерный путь $X: [a, b] \mapsto R$. Тогда сигнатура этого пути вычисляется следующим образом:

$$\begin{aligned} S(X)_{a,b}^1 &= X_b - X_a, \\ S(X)_{a,b}^{1,1} &= \frac{(X_b - X_a)^2}{2!}, \\ S(X)_{a,b}^{1,1,1} &= \frac{(X_b - X_a)^3}{3!}, \end{aligned}$$

...

$$S(X)_{a,b}^{1,1,\dots,1} = \frac{(X_b - X_a)^k}{k!}.$$

Отсюда можно видеть, что сигнатура по повторным индексам выражается через приращение по соответствующей координате, что верно и для многомерного случая.

Пример 2. Рассмотрим следующий двумерный путь $X: [a, b] \mapsto R^2$:

$$X_t = \{X_t^1, X_t^2\} = \{t, k \cdot t + d\},$$

$$dX_t = \{dX_t^1, dX_t^2\} = \{dt, k \cdot dt\},$$

где $k \neq 0$ – параметр. Тогда элементы сигнатуры $S(X)_{a,b}^{1,2}$ и $S(X)_{a,b}^{2,1}$ этого пути вычисляются следующим образом:

$$S(X)_{a,b}^{1,2} = \int_a^b \left(\int_a^{t_2} dt_1 \right) k dt_2 = \frac{k \cdot (b - a)^2}{2},$$

$$S(X)_{a,b}^{2,1} = \int_a^b \left(\int_a^{t_2} k dt_1 \right) dt_2 = \frac{k \cdot (b - a)^2}{2}.$$

Шафл-произведение

Одной из важных алгебраических особенностей сигнатуры является то, что произведение двух ее элементов $S(X)_{a,b}^{i_1, \dots, i_k}$ и $S(X)_{a,b}^{j_1, \dots, j_m}$ всегда может быть выражено через сумму других ее элементов, которая зависит исключительно от мультииндексов $(i_1, \dots, i_k)$ и $(j_1, \dots, j_m)$. Это свойство влечет за собой важнейшее следствие, показывающее отсутствие алгебраической независимости между членами сигнатуры. Кроме того, оно позволяет отказаться от непосредственного манипулирования произведениями элементов сигнатуры и перейти к работе с линейными комбинациями, что, в свою очередь, существенно упрощает их аналитическую обработку.

Далее введем понятие шафл-произведения двух мультииндексов.

Определение. Перестановка σ множества $\{1, \dots, k + m\}$ называется (k, m) -шафлом, если $\sigma^{-1}(1) < \dots < \sigma^{-1}(k)$ и $\sigma^{-1}(k + 1) < \dots < \sigma^{-1}(k + m)$.

Для множества всех (k, m) -шафлов будем использовать обозначение $\text{Shuffles}(k, m)$.

Рассмотрим два мульти-индекса $I = (i_1, \dots, i_k)$ и $J = (j_1, \dots, j_m)$, $i_1, \dots, i_k, j_1, \dots, j_m \in \{1, \dots, d\}$. Определим мультииндекс

$$(r_1, \dots, r_k, r_{k+1}, \dots, r_{k+m}) = (i_1, \dots, i_k, j_1, \dots, j_m).$$

Шафл-произведением I и J (обозначение: $I \# J$) называется конечный набор мульти-индексов длины $k+m$ вида

$$I \# J = \{(r_{\sigma(1)}, \dots, r_{\sigma(k+m)}) \mid \sigma \in \text{Shuffles}(k, m)\}.$$

Теорема. Для любого пути $X: [a, b] \mapsto R^d$ и мультииндексов $I = (i_1, \dots, i_k)$ и $J = (j_1, \dots, j_m)$, $i_1, \dots, i_k, j_1, \dots, j_m \in \{1, \dots, d\}$ верно равенство

$$S(X)_{a,b}^I S(X)_{a,b}^J = \sum_{K \in I \# J} S(X)_{a,b}^K.$$

Пример 3. Произведение элементов сигнатуры первого порядка выражается через элементы второго порядка следующим образом:

$$S(X)_{a,b}^1 S(X)_{a,b}^2 = S(X)_{a,b}^{1,2} + S(X)_{a,b}^{2,1}.$$

Геометрический смысл сигнатур

Независимость от начальной точки. Рассмотрим путь $X: [a, b] \mapsto R^d$ и $h \in R^d$. Пусть путь $Y: [x, y] \mapsto R^d$ имеет вид $Y_t = X_t + h$. Тогда

$$S(X)_{a,b} = S(Y)_{a,b}.$$

Независимость от репараметризации времени. Рассмотрим путь $X: [a, b] \mapsto R^d$ и биективную непрерывную неубывающую функцию $\psi: [x, y] \mapsto [a, b]$. Пусть путь $Y: [x, y] \mapsto R^d$ имеет вид $Y_t = X_{\psi_t}$. Тогда

$$S(X)_{a,b} = S(Y)_{x,y}.$$

Отметим, что первый уровень сигнатуры $(S(X)_{a,b}^1, \dots, S(X)_{a,b}^d)$, как видно из ее формального определения, совпадает с приращениями пути по отдельным координатам, в то время как второй уровень интерпретируется через понятие площади Леви (см., например, [3, 10]) – геометрической характеристики, указывающей на ориентированную площадь, охватываемую траекторией в двумерном пространстве.

Пусть $X: [a, b] \mapsto R^2$ – двумерный путь, где $X_t = (X_t^1, X_t^2)$. Проведем прямую от начальной до конечной точки пути, после чего все получившиеся площади при пересечении этой прямой и пути рассмотрим со знаком минус, если они выше прямой, и со знаком плюс, если они ниже прямой. **Площадью Леви** называется сумма данных площадей с учетом знаков, см. рис. 1.

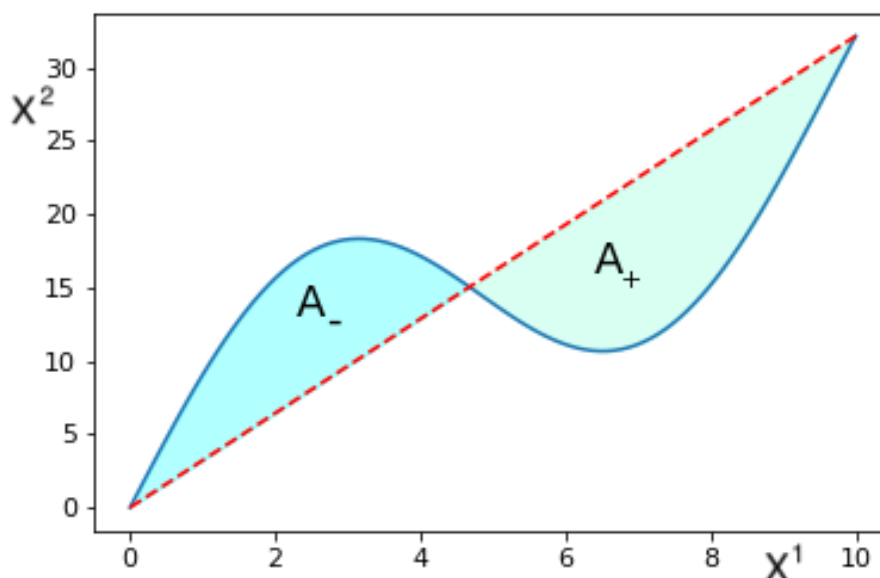


Рис. 1. Площадь Леви $A=A_+A_-$

Площадь Леви выражается через элементы сигнатуры следующим образом:

$$A = \frac{1}{2} (S(X)_{a,b}^{1,2} - S(X)_{a,b}^{2,1}).$$

Тождество Чена

Эффективность вычисления сигнатур в значительной мере зависит от природы исходного пути. В тех случаях, когда имеется аналитически заданная параметризация, соответствующие повторные интегралы могут быть получены в явной форме. Однако в большинстве прикладных задач, где траектории представляют собой последовательности наблюдений, их форма носит кусочно-линейный характер, а параметризация затруднена или вовсе отсутствует. В подобных случаях оказывается целесообразным применение тождества Чена, позволяющего рекурсивно вычислять сигнатуру всего пути по сигнатурам его составных отрезков.

Тождество Чена. Пусть $a < b < c$ и $X: [a, c] \mapsto R^d$. Тогда для любых $i_1, \dots, i_k \in W$ выполнено равенство

$$S(X)_{a,c}^{i_1, \dots, i_k} = \sum_{m=0}^k S(X)_{a,b}^{i_1, \dots, i_m} S(X)_{b,c}^{i_{m+1}, \dots, i_k}.$$

Для каждого одномерного сегмента сигнатура может быть получена напрямую, после чего производится поочередное соединение с предыдущими с использованием вышеуказанного тождества.

Пример 4. Рассмотрим путь $X: [0, 2] \mapsto R^2$ следующего вида, рис. 2:

$$X_t = \{t, 2 \cdot t\}, \quad t \in [0, 1],$$

$$X_t = \{t, 3 - t\}, \quad t \in [1, 2].$$

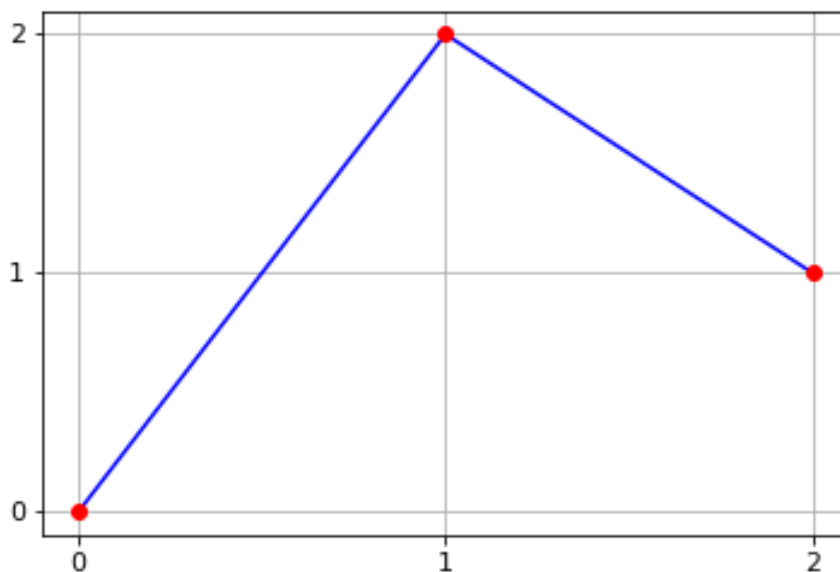


Рис. 2. Пример кусочно-линейного пути

Найдем первые два уровня сигнатуры пути отдельно на каждом из отрезков $[0, 1]$, $[1, 2]$. На этих отрезках путь является одномерным и линейным, поэтому, согласно примерам 1 и 2, первые два уровня сигнатуры этих участков выражаются следующим образом.

На отрезке [0,1]:

$$S(X)_{0,1}^{\emptyset} = 1 \text{ (по определению),}$$

$$S(X)_{0,1}^1 = \Delta X_{0,1}^1 = 1,$$

$$S(X)_{0,1}^{1,1} = \frac{(\Delta X_{0,1}^1)^2}{2} = 0.5,$$

$$S(X)_{0,1}^2 = \Delta X_{0,1}^2 = 2,$$

$$S(X)_{0,1}^{2,2} = \frac{(\Delta X_{0,1}^2)^2}{2} = 2,$$

$$S(X)_{0,1}^{1,2} = \frac{2 \cdot (1 - 0)^2}{2} = 1,$$

$$S(X)_{0,1}^{2,1} = \frac{2 \cdot (1 - 0)^2}{2} = 1.$$

На отрезке [1,2]:

$$S(X)_{1,2}^{\emptyset} = 1 \text{ (по определению),}$$

$$S(X)_{1,2}^1 = \Delta X_{1,2}^1 = 1,$$

$$S(X)_{1,2}^{1,1} = \frac{(\Delta X_{1,2}^1)^2}{2} = 0.5,$$

$$S(X)_{1,2}^2 = \Delta X_{1,2}^2 = -1,$$

$$S(X)_{1,2}^{2,2} = \frac{(\Delta X_{1,2}^2)^2}{2} = 0.5,$$

$$S(X)_{1,2}^{1,2} = \frac{-1 \cdot (2 - 1)^2}{2} = -0.5,$$

$$S(X)_{1,2}^{2,1} = \frac{-1 \cdot (2 - 1)^2}{2} = -0.5.$$

Применим тождество Чена:

$$S(X)_{0,2}^1 = S(X)_{0,1}^1 \cdot S(X)_{1,2}^{\emptyset} + S(X)_{0,1}^{\emptyset} \cdot S(X)_{1,2}^1 = 1 \cdot 1 + 1 \cdot 1 = 2 = \Delta X_{0,2}^1,$$

$$S(X)_{0,2}^2 = S(X)_{0,1}^2 \cdot S(X)_{1,2}^{\emptyset} + S(X)_{0,1}^{\emptyset} \cdot S(X)_{1,2}^2 = 2 \cdot 1 + 1 \cdot (-1) = 1 = \Delta X_{0,2}^2,$$

$$\begin{aligned} S(X)_{0,2}^{1,2} &= S(X)_{0,1}^{1,2} \cdot S(X)_{1,2}^{\emptyset} + S(X)_{0,1}^1 \cdot S(X)_{1,2}^2 + S(X)_{0,1}^{\emptyset} \cdot S(X)_{1,2}^{1,2} \\ &= 1 \cdot 1 + 1 \cdot (-1) + 1 \cdot (-0.5) = -0.5, \end{aligned}$$

$$S(X)_{0,2}^{2,1} = S(X)_{0,1}^{2,1} + S(X)_{0,1}^2 \cdot S(X)_{1,2}^1 + S(X)_{1,2}^{2,1} = 1 + 2 \cdot 1 - 0.5 = 2.5,$$

$$S(X)_{0,2}^{1,1} = S(X)_{0,1}^{1,1} + S(X)_{0,1}^1 \cdot S(X)_{1,2}^1 + S(X)_{1,2}^{1,1} = 0.5 + 1 \cdot 1 + 0.5 = 2 = \frac{(\Delta X_{0,2}^1)^2}{2},$$

$$S(X)_{0,2}^{2,2} = S(X)_{0,1}^{2,2} + S(X)_{0,1}^2 \cdot S(X)_{1,2}^2 + S(X)_{1,2}^{2,2} = 2 + 2 \cdot (-1) + 0.5 = 0.5$$

$$= \frac{(\Delta X_{0,2}^2)^2}{2}.$$

Логарифмические сигнатуры

Из теоремы о шафл-произведении вытекает, что элементы сигнатуры не являются алгебраически независимыми, однако для многих моделей машинного обучения требуются независимые входные данные для корректной работы и лучшего качества. Эту проблему решает понятие логарифмических сигнатур, которые, в отличие от обычных сигнатур, представляют собой алгебраически независимый набор величин, подходящих для прямого использования в качестве признаков пространства моделей.

Для определения понятия логарифмической сигнатуры введем сначала алгебру формальных степенных рядов.

Рассмотрим d формальных неопределенных величин $e_1, \dots, e_d$. Алгеброй некоммутативных формальных степенных рядов из d неопределенных называется векторное пространство всех рядов вида

$$\sum_{k=0}^{\infty} \sum_{i_1, \dots, i_k=1}^d \lambda_{i_1, \dots, i_k} e_{i_1} \dots e_{i_k},$$

где параметр второй суммы пробегает по всем мультииндексам $(i_1, \dots, i_k)$, $i_1, \dots, i_k \in 1, \dots, d$, и коэффициенты $\lambda_{i_1, \dots, i_k}$ являются вещественными числами.

На этом пространстве определены стандартные операции сложения рядов и умножения ряда на коэффициент, а также операция умножения рядов $\otimes$. При этом элементы пространства некоммутативны, т. е., например, $e_1 e_2$ и $e_2 e_1$ являются разными элементами.

Сигнатура может быть «закодирована» как элемент этого пространства (т. е. существует взаимно однозначное соответствие между сигнатурами и формальными степенными рядами):

$$S(X)_{a,b} = \sum_{k=0}^{\infty} \sum_{i_1, \dots, i_k=1}^d S(X)_{a,b}^{i_1, \dots, i_k} e_{i_1} \dots e_{i_k}.$$

Иными словами, элементы сигнатуры рассматриваются как соответствующие коэффициенты ряда, что позволяет «перейти» в алгебру рядов. В ней можно выполнять операции сложения и умножения рядов, а коэффициенты полученного ряда интерпретировать как элементы сигнатуры.

Далее для формального ряда, в котором первый коэффициент равен 1, а остальные равны 0 (что соответствует сигнатуре $\{1, 0, 0, \dots\}$), будем использовать обозначение 1.

Логарифмической сигнатурой пути $X: [a, b] \mapsto R^d$ называется формальный степенной ряд

$$\log S(X)_{a,b} = \sum_{n=1}^{\infty} \frac{(-1)^n}{n} (1 - S(X)_{a,b})^{\otimes n}.$$

Тем самым, логарифмическая сигнатура становится инструментом генерации независимого набора признаков, обладающих полной описательной способностью по отношению к геометрии исходного многомерного пути.

Следует, однако, указать, что сигнатура не позволяет восстановить траекторию в полном объеме. Так, например, информация о скорости прохождения пути утрачивается в силу инвариантности сигнатуры относительно репараметризации времени. Тем не менее при соблюдении определенных условий, в частности при отсутствии самопересечений, сигнатура полностью определяет как множество точек, через которые проходит путь, так и порядок их обхода, что открывает возможности для применения в задачах восстановления структуры данных.

РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЙ

Практика предоставляет множество различных задач, в которых неизвестны распределение и стохастическая основа данных, в связи с чем нет возможности установить математические свойства модели. Модели, описывающие эти задачи, можно значительно усложнять, подбирая и добавляя в них новые параметры, однако в реальности они не смогут полностью описать наблюдаемое поведение.

Помимо этого, при использовании временных рядов во многих ситуациях может оказаться недостаточным количество данных для применения моделей глубокого машинного обучения, а также данные могут быть конфиденциальны и недоступны для обучения моделей. Примерами этого могут быть образовательные, медицинские и финансовые данные.

В ситуациях, когда необходимо расширить доступное множество данных, не прибегая к гипотезам об их вероятностной природе, можно использовать сигнатурные методы в сочетании с генеративными моделями. Так, на основе оригинального множества временных рядов, представленных в сигнатурной форме, можно синтезировать новые данные, демонстрирующие похожее распределение. При этом необходимость знания истинного распределения отпадает, поскольку оценка схожести может быть проведена с помощью, например, метрики максимального среднего расхождения (MMD), построенной также на основе сигнатур [16].

Данные образовательного цифрового следа студентов являются ярким примером временных рядов для которых неизвестны их распределение и стохастическая основа. Кроме того, в зависимости от университета, курса и среднего уровня студентов они могут отличаться, что вводит еще большие ограничения на количество доступных данных. Поэтому для тестирования различных педагогических гипотез, а также для создания моделей по подбору персональных траекторий и построению поведенческой и предсказательной аналитики необходимы многолетние сборы студенческих данных.

В связи с этим были применены методы генерации, анализа и прогноза траекторий обучения студентов, использующие сигнатуры исходного набора данных. Преимуществом использования сигнатурных методов является то, что они способны улавливать взаимосвязи между различными размерностями временного ряда. Например, появляется возможность обнаруживать взаимные зависимости между посещаемостью, временем, потраченным на решение задач, плагиатом, долей набранных за занятия баллов и другими метриками цифрового следа студентов. Это значительно повышает общее качество модели, поскольку рассмотрение отдельных характеристик независимо друг от друга не дает полной картины, например, у студента могут быть высокие баллы, но при высоком уровне

плагиата, что изменяет его общий статус относительно группы. Помимо этого, у модели появляется возможность улавливать, например, повышение плагиата при сокращении времени решения задач, снижение среднего балла при падении посещаемости и другие зависимости.

Такой подход, реализованный в контексте образовательной аналитики, позволяет на массиве данных, синтетически сгенерированных с помощью сигнатурных методов, построить модель прогнозирования поведения студента, в которой сигнатуры временных рядов выступают в роли признакового пространства. Эта модель, в свою очередь, служит инструментом визуализации и интерпретации динамики учебных результатов, указывая преподавателю на необходимость индивидуального вмешательства или корректировки образовательной траектории с целью максимизации обучающего эффекта. Более того, на основе такого прогностического механизма можно автоматически сформировать персонализированные образовательные траектории, предложив учащимся релевантные задания, а также вспомогательные материалы, соответствующие их текущему уровню и способствующие увеличению эффективности их обучения.

На рис. 3, представлены результаты работы обученной прогностической модели, которая предсказывает следующее значение траектории цифрового следа студента на основе окна из 5 предыдущих значений. Вертикальной линией отмечен момент, с которого модель начинает свою работу, таким образом, уже спустя примерно месяц реального преподавания у модели появляется возможность предсказывать будущее поведение студента. При сдвиге окна каждый раз в качестве нового значения добавляется истинное значение. Такой подход хорошо соотносится с реальным педагогическим применением модели, так как она в режиме реального времени имеет доступ к актуальным данным студента на текущую дату.



Рис. 3. Пример работы модели предсказания траектории студента с помощью сигнатурных методов.

ЗАКЛЮЧЕНИЕ

Сигнатурные методы доказали свою эффективность при анализе временных рядов, предложив гибкий инструмент для машинного обучения и генерации данных. Они преодолевают ограничения традиционных подходов и показывают высокую точность в прогнозировании. Особую ценность сигнатурные методы представляют в условиях ограниченного объема данных, поскольку позволяют генерировать синтетические данные, сохраняющие распределение исходных. Это открывает новые возможности для применения в образовании, медицине, финансах и других областях, где данные часто являются конфиденциальными или труднодоступными. Результаты экспериментальных исследований, представленные в статье, подтверждают практическую значимость сигнатурных методов, особенно в анализе образовательных траекторий студентов. Использование этих методов позволяет не только прогнозировать поведение студентов, но и адаптировать учебные программы для достижения наилучших результатов.

Благодарности

Работа выполнена в рамках темы государственного задания НИЦ «Курчатовский институт» – НИИСИ по теме № FNEF-2024-0001 (1023032100070-3-1.2.1).

СПИСОК ЛИТЕРАТУРЫ

1. *Chen K.-T.* Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula // *Annals of Mathematics*. 1957. Vol. 65. No. 1. P. 163–178. <https://doi.org/10.2307/1969671>
2. *Chen K.-T.* Integration of paths – a faithful representation of paths by non-commutative formal power series // *Transactions of the American Mathematical Society*. 1958. Vol. 89. P. 395–407. <https://doi.org/10.1090/S0002-9947-1958-0106258-0>
3. *Lyons T.J.* Differential equations driven by rough signals // *Revista Matemática Iberoamericana*. 1998. Vol. 14. No. 2. P. 215–310. <https://doi.org/10.4171/RMI/240>
4. *Lyons T.J., Caruana M., Levy T.* Differential equations driven by rough paths // *Lecture Notes in Mathematics*. 2007. № 1908. <https://doi.org/10.1007/978-3-540-71285-5>
5. *Friz P.K., Hairer M.* A course on rough paths. With an introduction to regularity structures (2nd edition) // Springer. 2020. <https://doi.org/10.1007/978-3-030-41556-3>
6. *Boedihardjo H., Geng X., Lyons T., Yang D.* The signature of a rough path: Uniqueness. 2014. <https://arxiv.org/abs/1406.7871>
7. *Hambly B., Lyons T.* Uniqueness for the signature of a path of bounded variation and the reduced path group // *Annals of Mathematics*. 2010. Vol. 171. No. 1. P. 109–167. <https://doi.org/10.4007/annals.2010.171.109>
8. *Graham B.* Sparse arrays of signatures for online character recognition. 2013. <https://arxiv.org/abs/1308.0371>
9. *Morrill J., Kormilitzin A., Nevado-Holgado A., Swaminathan S., Howison S., Lyons T.J.* The signature-based model for early detection of sepsis from electronic health records in the intensive care unit // *IEEE Conference on Computing in Cardiology*. 2019. <https://doi.org/10.22489/CinC.2019.014>

10. Chevyrev I., Kormilitzin A. A Primer on the Signature Method in Machine Learning. 2019. <https://arxiv.org/abs/1603.03788>
 11. Buhler H., Horvath B., Lyons T., Arribas I. P., Wood B. A data-driven market simulator for small data environments. 2020. <https://arxiv.org/abs/2006.14498>
 12. Gyurko L.G., Lyons T., Kontkowski M., Field J. Extracting information from the signature of a financial data stream. 2014. <https://arxiv.org/abs/1307.7244v2>
 13. Cohen S.N., Lui S., Malpass W., Mantoan G., Nesheim L., de Paula A., Reeves A., Scott C., Small E., Yang L. Nowcasting with signature methods. 2023. <https://arxiv.org/abs/2305.10256v1>
 14. Zou H., Hastie T. Regularization and variable selection via the elastic net // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005. Vol. 67. No. 2. P. 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
 15. Fermanian A. Embedding and learning with signatures // Computational Statistics and Data Analysis. 2021. Vol. 157. No. 107148. <https://doi.org/10.1016/j.csda.2020.107148>
 16. Chevyrev I., Oberhauser H. Signature moments to characterize laws of stochastic processes. 2018. <https://arxiv.org/abs/1810.10971>
-

SIGNATURE METHODS FOR TIME SERIES ANALYSIS

K. A. Mashchenko^[0000-0003-0355-6699]

NRC “Kurchatov Institute” – SRISA, Moscow, 117218, Russia

kirill.mashchenko@niisi.ru

Abstract

Signature methods are a powerful tool for time series analysis, transforming them into a form suitable for machine learning tasks. The article examines the fundamental concepts of path signatures, their properties, and geometric interpretation, as well as computational methods for various types of time series. Examples of signature method applications in different fields, including finance, medicine, and education, are presented, highlighting their advantages over traditional approaches. Special attention is given to synthetic data generation based on signatures, which is particularly relevant

when working with limited datasets. The experimental results on generating and predicting student digital learning trajectories demonstrate the effectiveness of signature methods for machine learning applications in time series analysis and forecasting.

Keywords: *signature, signature methods, time series, data generation, trajectory analysis, digital footprint*

REFERENCES

1. *Chen K.-T.* Integration of paths, geometric invariants and a generalized Baker-Hausdorff formula // *Annals of Mathematics*. 1957. Vol. 65. No. 1. P. 163–178. <https://doi.org/10.2307/1969671>
2. *Chen K.-T.* Integration of paths – a faithful representation of paths by non-commutative formal power series // *Transactions of the American Mathematical Society*. 1958. Vol. 89. P. 395–407. <https://doi.org/10.1090/S0002-9947-1958-0106258-0>
3. *Lyons T.J.* Differential equations driven by rough signals // *Revista Matemática Iberoamericana*. 1998. Vol. 14. No. 2. P. 215–310. <https://doi.org/10.4171/RMI/240>
4. *Lyons T.J., Caruana M., Levy T.* Differential equations driven by rough paths // *Lecture Notes in Mathematics*. 2007. № 1908. <https://doi.org/10.1007/978-3-540-71285-5>
5. *Friz P.K., Hairer M.* A course on rough paths. With an introduction to regularity structures (2nd edition) // Springer. 2020. <https://doi.org/10.1007/978-3-030-41556-3>
6. *Boedihardjo H., Geng X., Lyons T., Yang D.* The signature of a rough path: Uniqueness. 2014. <https://arxiv.org/abs/1406.7871>
7. *Hambly B., Lyons T.* Uniqueness for the signature of a path of bounded variation and the reduced path group // *Annals of Mathematics*. 2010. Vol. 171. No. 1. P. 109–167. <https://doi.org/10.4007/annals.2010.171.109>
8. *Graham B.* Sparse arrays of signatures for online character recognition. 2013. <https://arxiv.org/abs/1308.0371>
9. *Morrill J., Kormilitzin A., Nevado-Holgado A., Swaminathan S., Howison S., Lyons T.J.* The signature-based model for early detection of sepsis from electronic health records in the intensive care unit // *IEEE Conference on Computing in Cardiology*. 2019. <https://doi.org/10.22489/CinC.2019.014>

10. *Chevyrev I., Kormilitzin A.* A Primer on the Signature Method in Machine Learning. 2019. <https://arxiv.org/abs/1603.03788>
11. *Buhler H., Horvath B., Lyons T., Arribas I. P., Wood B.* A data-driven market simulator for small data environments. 2020. <https://arxiv.org/abs/2006.14498>
12. *Gyurko L.G., Lyons T., Kontkowski M., Field J.* Extracting information from the signature of a financial data stream. 2014. <https://arxiv.org/abs/1307.7244v2>
13. *Cohen S.N., Lui S., Malpass W., Mantoan G., Nesheim L., de Paula A., Reeves A., Scott C., Small E., Yang L.* Nowcasting with signature methods. 2023. <https://arxiv.org/abs/2305.10256v1>
14. *Zou H., Hastie T.* Regularization and variable selection via the elastic net // Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2005. Vol. 67. No. 2. P. 301–320. <https://doi.org/10.1111/j.1467-9868.2005.00503.x>
15. *Fermanian A.* Embedding and learning with signatures // Computational Statistics and Data Analysis. 2021. Vol. 157. No. 107148. <https://doi.org/10.1016/j.csda.2020.107148>
16. *Chevyrev I., Oberhauser H.* Signature moments to characterize laws of stochastic processes. 2018. <https://arxiv.org/abs/1810.10971>

СВЕДЕНИЯ ОБ АВТОРЕ



МАЩЕНКО Кирилл Алексеевич. Младший научный сотрудник отдела учебной информатики НИЦ «Курчатовский институт» – НИИСИ, младший научный сотрудник Лаборатории вычислительных методов механико-математического факультета МГУ имени М.В. Ломоносова. Области исследований: автоматизация и цифровизация образовательного процесса, в том числе с применением методов искусственного интеллекта.

Kirill A. MASHCHENKO. Junior Researcher at the Department of Educational Informatics, NRC “Kurchatov Institute” – SRISA, Junior Researcher at the Laboratory of Computational Methods, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University. Research areas: Automation and digital transformation of the educational process, including the application of artificial intelligence methods.

email: kirill.mashchenko@niisi.ru

ORCID: 0000-0003-0355-6699

Материал поступил в редакцию 10 апреля 2025 года

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПОВЫШЕНИЯ КАЧЕСТВА ТЕСТОВ

Р. Р. Минюков¹ [0009-0008-1264-7107], М. М. Абрамский² [0000-0003-3063-8948]

^{1, 2}Казанский (Приволжский) федеральный университет, г. Казань, 420008, Россия

¹ramil.minyukov@gmail.com, ²mabramsk@kpfu.ru

Аннотация

Работа посвящена применению методов машинного обучения для повышения качества тестов. Проведен обзор предметной области и реализованы два метода повышения качества: поиск похожих вопросов и оценка качества дистракторов. Первый включает тестирование пяти моделей трансформеров для получения векторного представления текста и шесть алгоритмов кластеризации. Второй метод основан на использовании тех же моделей трансформеров совместно с тремя алгоритмами классификации. Результаты экспериментов показали высокую эффективность предложенных решений при решении обеих задач.

Ключевые слова: анализ тестовых вопросов, дистракторы, машинное обучение, прохождение тестов, тесты, повышение качества тестов.

ВВЕДЕНИЕ

Рядом исследователей поднимается вопрос выбора валидной подборки вопросов, которые входят в конкретный тест, а также в целом повышения качества тестовых заданий (см., например, [1–3]).

Особое внимание уделяется вопросам автоматического анализа тестовых заданий. Исследования [4, 5] акцентируют внимание на вопросах выбора релевантных вопросов, а также повышения общей надежности и валидности тестов.

В настоящем исследовании была поставлена цель разработки методов и программных средств, направленных на повышение качества тестирования за счет анализа и реализации моделей искусственного интеллекта (ИИ), применяемых при автоматизированном исправлении заданий.

Статья организована следующим образом:

- в первом разделе проведен анализ предметной области и дан обзор существующих моделей и подходов к применению моделей и технологий ИИ в оценке заданий;
- во втором разделе рассмотрены и реализованы методы поиска похожих вопросов на основе их семантической близости;
- в третьем разделе речь рассмотрена реализация метода определения качества дистракторов в заданиях со множественным выбором.

1. СУЩЕСТВУЮЩИЕ ПОДХОДЫ К ПОВЫШЕНИЮ КАЧЕСТВА ТЕСТОВ

На данный момент существует несколько подходов к анализу тестовых заданий. Широко используемый из них основан на статистических показателях [5–7], к которым можно отнести:

- индекс легкости тестового задания;
- среднеквадратичное отклонение;
- индекс дифференциации;
- эффективность дистракторов;
- коэффициент угадывания;
- оценка надежности по методу расщепления на эквивалентные половины.

Статистические параметры не учитывают смысловую часть заданий, поэтому необходимо использовать технологии ИИ для нахождения семантических связей между элементами теста.

Одним из способов повышения качества заданий за счет методов ИИ является поиск похожих вопросов, снижающих эффективность теста, поскольку они могут дублировать проверяемые знания, создавая лишнюю когнитивную нагрузку [8]. Одним из подходов является кластеризация тестовых вопросов, которая обсуждается, например, в работах [9–11].

В [12] затронуты вопросы анализа качества дистракторов. Известные данные показывают, что опытные преподаватели редко создают более двух валидных дистракторов. Это подтверждает необходимость регулярного анализа и удаления неэффективных отвлекающих вариантов.

Еще одним из подходов является автоматический поиск скрытых подсказок, содержащихся в формулировках вопросов или вариантах ответов, при котором один вопрос может содержать информацию для ответа на другой [13]. Автоматизация такого подхода значительно повышает эффективность проверки и повышения качества тестовых заданий.

В результате обзора исследований, посвященных использованию методов и подходов искусственного интеллекта, выделены алгоритмы, которые способствуют повышению качества академических тестов.

2. ПОИСК ПОХОЖИХ ВОПРОСОВ

Для сравнения различных алгоритмов определения похожих заданий требуется предварительная подготовка данных.

Для английского языка известен широко используемый набор данных Quora Question Pairs [14]. Он содержит пары вопросов с метками, указывающими на то, являются ли вопросы дубликатами. Этот набор данных применяется в задачах определения семантической близости и поиска похожих вопросов.

Для русского языка имеется аналогичный по структуре набор данных Quora question pairs russian [15]. Он также содержит пары предложений и бинарную метку схожести.

Для тестирования алгоритмов данные из обоих наборов были предварительно агрегированы в группы фиксированного размера. В рамках проведенного эксперимента каждая группа состояла из 10 вопросов, внутри которых в среднем содержалось от 2 до 6 схожих по смыслу формулировок. Такая структура позволяла оценивать способность моделей находить дубликаты в ограниченном контексте.

Для удобства обработки данных и их подготовки к обучению разработаны вспомогательные классы, обеспечивающие удобное обращение с данными. Эти классы содержат методы, которые возвращают следующие данные:

- уникальные идентификаторы вопросов;
- список вопросов;
- список групп вопросов, схожих между собой.

В рамках проведенного исследования для получения векторного представления вопросов были выбраны модели Sentence Transformers, взятые с платформы Hugging Face [16]:

- all-MiniLM-L6-v2 [17];
- sentence-transformers/all-mpnet-base-v2 [18];
- sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 [19];
- Alibaba-NLP/gte-multilingual-base [20];
- cointegrated/rubert-tiny2 [21].

Выбор этих моделей обусловлен их различной архитектурой, размерностью и способностью обрабатывать тексты на нескольких языках. Некоторые из них, например sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 и Alibaba-NLP/gte-multilingual-base, обучены на многоязычных наборах данных и способны сохранять семантические связи вне зависимости от языка входной последовательности. Это особенно полезно для задач на русском и английском языках одновременно.

В качестве алгоритмов и методов кластеризации были выбраны следующие алгоритмы:

- косинусное сходство с пороговыми значениями 0.7, 0.8 и 0.9;
- К-средних (K-Means);
- иерархическая кластеризация (Agglomerative Clustering);
- гауссовы смеси (Gaussian Mixture Models);
- метод Санору в сочетании с алгоритмом К-средних;
- разделяющий (бисекционный) алгоритм К-средних (Bisecting K-Means).

Для количественной оценки качества были выбраны следующие метрики: Adjusted Rand Index (ARI) и нормализованная взаимная информация (NMI).

Результаты тестирования представлены в табл. 1 и 2. Наилучшие результаты показали следующие комбинации:

- Alibaba-NLP/gte-multilingual-base и косинусное сходство;
- Alibaba-NLP/gte-multilingual-base и К-средних;
- Alibaba-NLP/gte-multilingual-base и иерархическая кластеризация;
- Alibaba-NLP/gte-multilingual-base и гауссовы смеси.

Таблица 1. Результаты тестирования моделей Sentence Transformers (по языку тестирования и порогу схожести)

Модель Sentence Transformers		all-MiniLM-L6-v2		all-mpnet-base-v2		gte-multilingual-base		paraphrase-multilingual MiniLM-L12-v2		rubert-tiny2	
	Язык	Англ.	Рус.	Англ.	Рус.	Англ.	Рус.	Англ.	Рус.	Англ.	Рус.
Порог схожести	Метрика тестирования										
0.70	Скорректированный индекс Рэнда (ARI)	0.68	0.51	0.72	0.51	0.71	0.64	0.68	0.59	0.44	0.56
0.80	Скорректированный индекс Рэнда (ARI)	0.70	0.49	0.73	0.51	0.71	0.64	0.71	0.59	0.44	0.56
0.90	Скорректированный индекс Рэнда (ARI)	0.67	0.49	0.71	0.52	0.72	0.65	0.69	0.61	0.46	0.57
0.70	Нормализованная взаимная информация (NMI)	0.97	0.95	0.97	0.95	0.97	0.97	0.97	0.96	0.90	0.96
0.80	Нормализованная взаимная информация (NMI)	0.97	0.95	0.98	0.95	0.97	0.97	0.97	0.97	0.90	0.96
0.90	Нормализованная взаимная информация (NMI)	0.97	0.95	0.97	0.95	0.97	0.97	0.97	0.97	0.91	0.96

Таблица. 2. Результаты тестирования моделей Sentence Transformers в зависимости от языка тестирования и алгоритма кластеризации

Модель Sentence Transformers		all-MiniLM-L6-v2		all-mpnet-base-v2		gte-multilingual-base		paraphrase-multilingual-MiniLM-L12-v2		rubert-tiny2	
	Язык	Англ.	Рус.	Англ.	Рус.	Англ.	Рус.	Англ.	Рус.	Англ.	Рус.
Скорректированный индекс Рэнда (ARI)	Bisecting K-средних	0.04	0.03	0.03	0.03	0.04	0.03	0.03	0.03	0.04	0.03
	K-средних	0.53	0.42	0.53	0.40	0.53	0.52	0.52	0.48	0.51	0.49
	Сапору + K-средних	0.61	0.02	0.62	0.02	0.34	0.28	0.09	0.20	0.00	0.07
	Гауссовы смеси	0.53	0.42	0.53	0.40	0.53	0.52	0.52	0.48	0.51	0.49
	Иерархическая кластеризация	0.53	0.39	0.52	0.36	0.52	0.50	0.51	0.46	0.50	0.47
Нормализованная взаимная информация (NMI)	Bisecting K-средних	0.41	0.39	0.41	0.38	0.42	0.40	0.40	0.37	0.40	0.40
	K-средних	0.92	0.89	0.92	0.87	0.92	0.92	0.92	0.90	0.92	0.92
	Сапору + K-средних	0.95	0.36	0.95	0.40	0.87	0.85	0.96	0.96	0.01	0.66
	Гауссовы смеси	0.92	0.89	0.92	0.87	0.92	0.92	0.92	0.90	0.92	0.92
	Иерархическая кластеризация	0.92	0.85	0.92	0.79	0.92	0.92	0.91	0.89	0.92	0.91

3. РЕАЛИЗАЦИЯ МЕТОДА ОЦЕНКИ ДИСТРАКТОРОВ

Для реализации метода, предназначенного для определения эффективности дистракторов, был выбран набор данных «Многоязычное измерение массового многозадачного понимания языка» (Multilingual Measuring Massive Multitask Language Understanding, MMLU) [22]. Его выбор обусловлен следующими факторами.

Во-первых, этот набор содержит более 50 тем, что обеспечивает разнообразие вопросов и их тематик. Во-вторых, набор поддерживает мультиязычность: вопросы и ответы переведены сразу на несколько языков, в том числе на русский. Это позволяет обучать модель сразу на нескольких языках. В-третьих, MMLU имеет удобную структуру заданий: каждый вопрос оформлен в виде множественного выбора с одним правильным ответом и несколькими дистракторами.

Для создания метода было необходимо сгенерировать отвлекающие варианты ответов: был разработан специализированный алгоритм их автоматизированного построения, применимый для русского и английского языков. Отличался только процесс выбора синонимов, где использовались различные лексические базы слов.

Общий алгоритм заключается в следующем. На первом этапе определялся тип правильного ответа: одиночное слово, числовое значение или короткое выражение. В случае, когда правильным ответом является слово, для генерации использованы его синонимы, случайные числовые значения, а также отдельные слова, случайно выбранные из вопроса. Если ответ представляет собой число, то в его качестве подбираются значения, гораздо большие или меньшие исходного значения, либо случайные ответы из общего набора ответов. В ситуации, когда правильным ответом является выражение, дистракторы формируются за счет выбора случайных ответов из общего набора.

Различие в построении отвлекающих вопросов между русским и английским языками заключается исключительно в этапе поиска синонимов. Для русского языка синонимы извлекались из лексической базы wiki_ru_wordnet [23], тогда как для английского использовалась база WordNet, доступная через библиотеку NLTK [24]. Все остальные этапы являются идентичными для разных языков, что обеспечивает единообразие подхода при работе с многоязычными данными.

Для подготовки данных к обучению были использованы пять различных моделей получения векторного представления слов, аналогичные поиску похожих вопросов. Каждая модель применялась для вычисления косинусного сходства для пар «вопрос – дистрактор» и «правильный ответ – дистрактор». Полученные метрики позволяют количественно оценить, насколько семантически близки элементы задания.

Для каждого дистрактора рассчитывались следующие показатели:

- косинусное сходство между векторными представлениями дистрактора и правильного ответа;
- косинусное сходство между векторными представлениями дистрактора и формулировкой вопроса;
- длина дистрактора, измеряемая количеством слов;
- разность в длине между дистрактором и правильным ответом.

На этапе обучения были протестированы три модели бинарной классификации: логистическая регрессия (Logistic Regression), метод опорных векторов (Support Vector Classifier, SVC) и метод случайного леса (Random Forest). Выбор моделей обусловлен их широкой распространенностью в задачах бинарной классификации.

Использование названных моделей обосновано их применением в аналогичных задачах анализа качества дистракторов. Например, логистическая регрессия и метод случайного леса применялись в работе [25] в качестве базового классификатора и метода ранжирования потенциальных дистракторов.

Для всех моделей зафиксированы следующие параметры:

- Logistic Regression – параметры по умолчанию;
- SVC – использовались StandardScaler и gamma, равная auto;
- Random Forest – параметр n_estimators, равный 100.

Для оценки качества моделей использованы следующие метрики:

- Accuracy – доля правильно классифицированных примеров;
- Macro-F1 – среднее гармоническое значение precision и recall по всем классам без учета их веса;
- Weighted-F1 – взвешенное среднее F1-score, учитывающее несбалансированность классов.

Результаты тестирования моделей представлены на рис. 1–3.

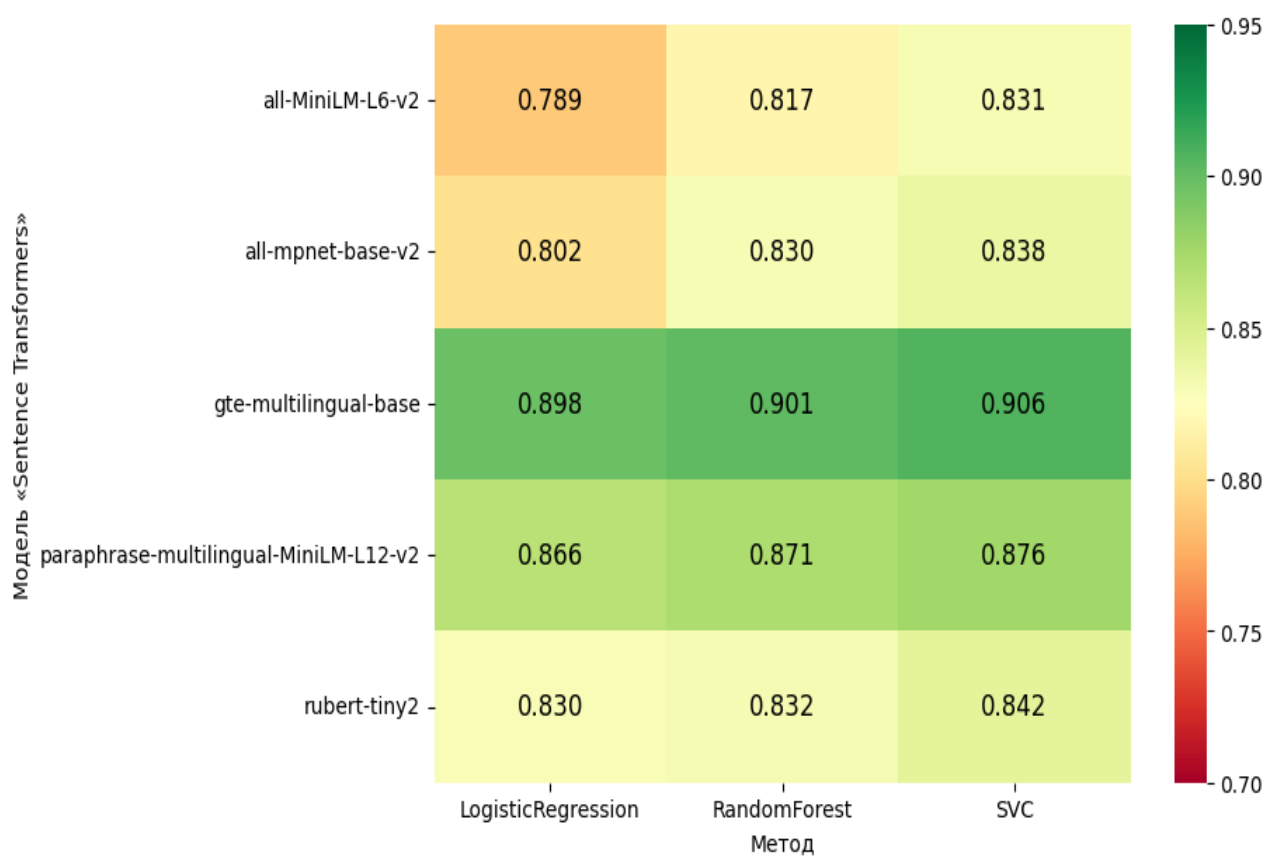


Рис. 1. Сравнение моделей Sentence Transformers по метрике Accuracy для методов классификации

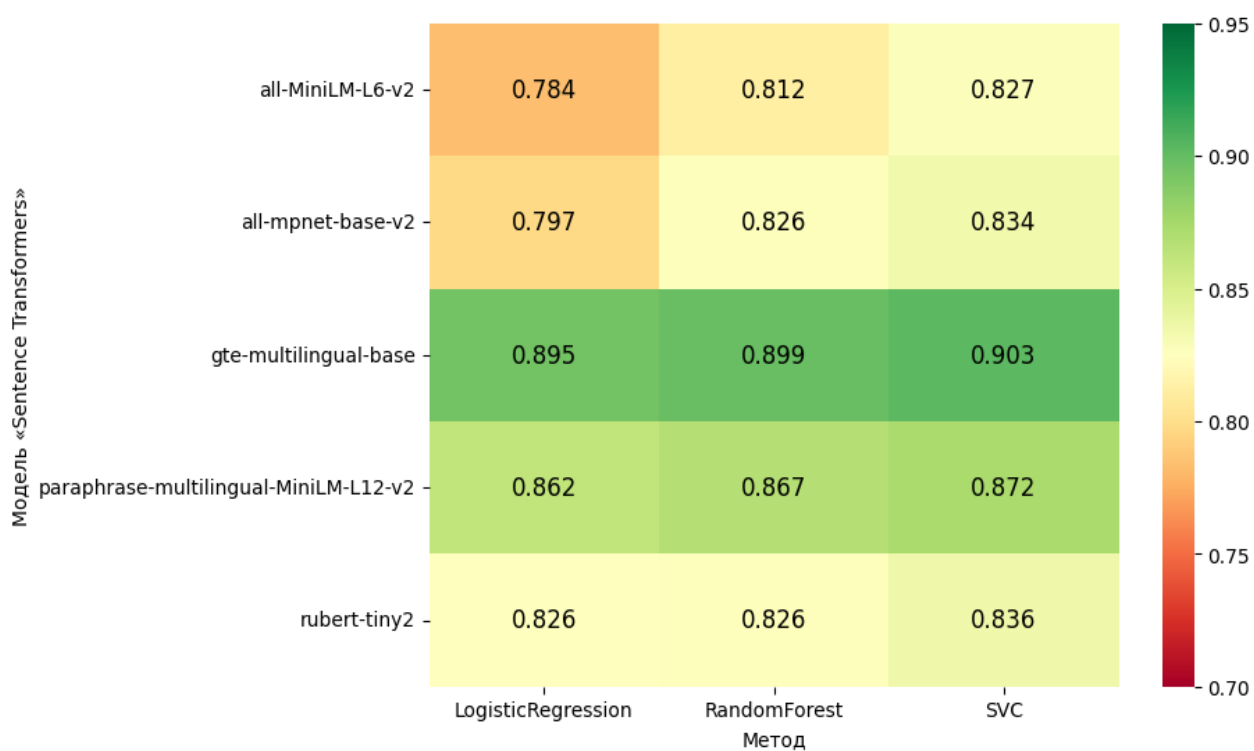


Рис. 2. Сравнение моделей Sentence Transformers по метрике Macro-F1 для методов классификации

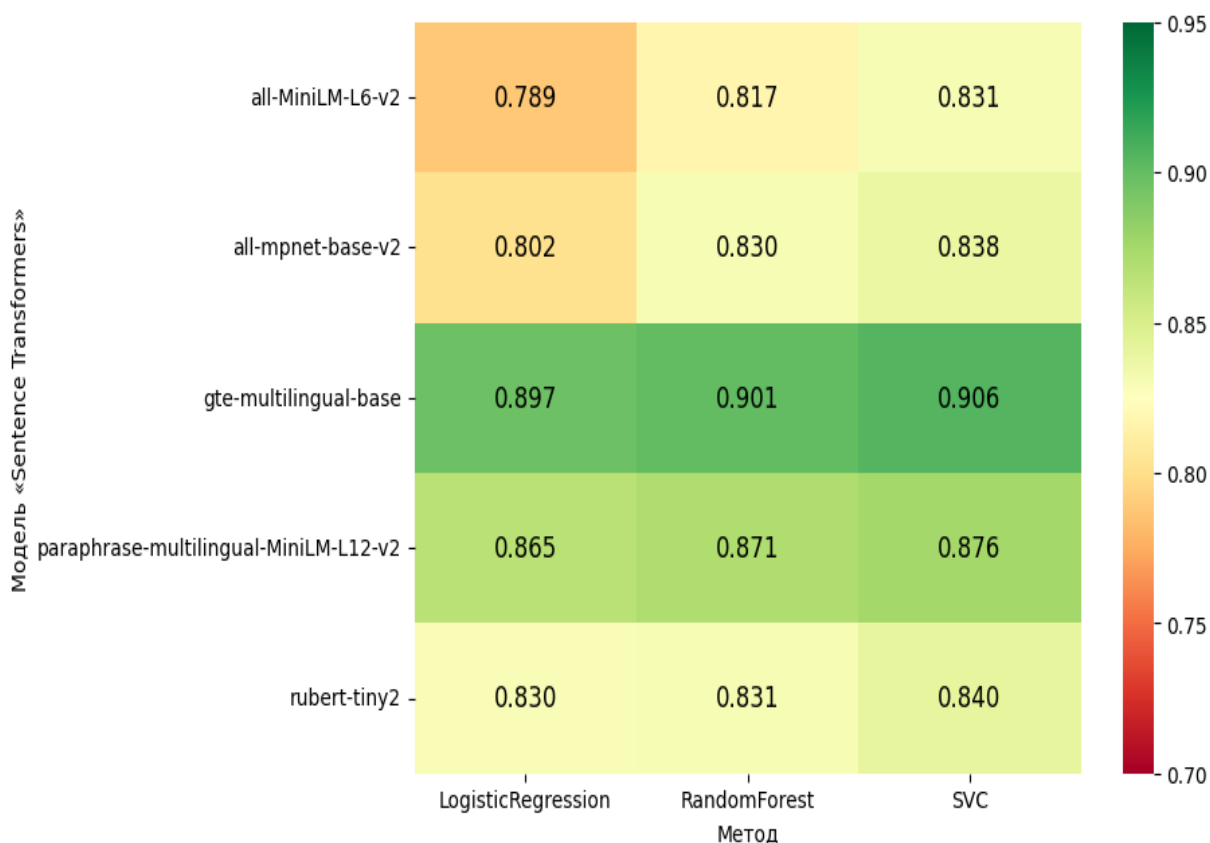


Рис. 3. Сравнение моделей Sentence Transformers по метрике Weighted-F1 для методов классификации

Результаты показали, что наилучших показателей удалось достичь, используя модель опорных векторов в сочетании с моделью получения эмбеддингов Alibaba-NLP/gte-multilingual-base.

ЗАКЛЮЧЕНИЕ

Таким образом, в ходе экспериментов установлено, что сочетания моделей Sentence Transformers и алгоритмов кластеризации способны эффективно находить семантические связи в формулировках вопросов, а сочетание Sentence Transformers и метода опорных векторов – определять качество дистракторов. Это позволяет сократить количество ручных проверок и автоматизировать процесс повышения качества тестов.

Дальнейшая работа направлена на разработку метода определения подсказок в формулировках вопросов и вариантах ответов.

СПИСОК ЛИТЕРАТУРЫ

1. Чельшкова М.Б. Теория и практика конструирования педагогических тестов. М.: Логос, 2002. Т. 432. С. 5.
2. Hrich N., Azekri M., Khaldi M. An ai educational tool for detecting redundancy in distractors and items within multiple-choice tests // INTED2024 Proceedings. IATED, 2024. P. 6454–6458.
3. Аванесов В.С. Теория и практика педагогических измерений. ЦТ и МКО УГТУ-УПИ, 2005.
4. Brusilovsky P., Miller P. Web-based testing for distance education. In: P. DeBra and J. Leggett (Eds.) Proceedings of WebNet'99, World Conference of the WWW and Internet, Honolulu, HI, Oct. 24–30, 1999. AACE, P. 149–154.
5. Толстобров А.П., Коржик И.А. Возможности анализа и повышения качества тестовых заданий при использовании сетевой системы управления обучением Moodle // Вестник ВГУ. 2008. Т. 2. С. 100–106.
6. Алпацкая Е.В., Бубнов Н.В., Минченков А.В. Дифференцирующая способность тестовых материалов для оценки качества обучения // Ученые записки университета им. П.Ф. Лесгафта. 2015. № 11 (129). С. 9–14.
7. Анастаси А. Психологическое тестирование. Питер, 2009.
8. Lord F.M. Applications of item response theory to practical testing problems. Routledge, 2012.
9. AlMahmoud R.H., Alian M. The effect of clustering algorithms on question answering // Expert Systems with Applications. 2024. Vol. 243. Article number 122959.
10. Zhang W.N. et al. A topic clustering approach to finding similar questions from large question and answer archives // PloS one. 2014. Vol. 9, No. 3. Article number e71511.
11. Alian M., Al-Naymat G. Questions clustering using canopy-K-means and hierarchical-K-means clustering // International Journal of Information Technology, 2022. Vol. 14, No. 7. P. 3793–3802.
12. Tarrant M., Ware J., Mohammed A. M. An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis // BMC medical education. 2009. Vol. 9. P. 1–8.

13. Moore S. et al. An automatic question usability evaluation toolkit // International Conference on Artificial Intelligence in Education. Cham: Springer Nature Switzerland, 2024. P. 31–46.
14. First Quora Dataset Release: Question Pairs // Quora, 2017.
URL: <https://www.quora.com/q/quoradata/First-Quora-Dataset-Release-Question-Pairs> (дата обращения: 30.04.2025).
15. Quora Question Pairs Russian // Платформа Kaggle, 2022.
URL: <https://www.kaggle.com/datasets/loopdigga/quora-question-pairs-russian> (дата обращения: 30.04.2025).
16. Hugging Face – The AI Community [Электронный ресурс].
URL: <https://huggingface.co/> (дата обращения: 30.04.2025).
17. All-MiniLM-L6-v2 // Платформа Hugging Face.
URL: <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2> (дата обращения: 30.04.2025).
18. Sentence-transformers/all-mpnet-base-v2 // Платформа Hugging Face.
URL: <https://huggingface.co/sentence-transformers/all-mpnet-base-v2> (дата обращения: 30.04.2025).
19. Paraphrase-multilingual-MiniLM-L12-v2 // Платформа Hugging Face.
URL: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> (дата обращения: 30.04.2025).
20. Gte-multilingual-base // Платформа Hugging Face.
URL: <https://huggingface.co/Alibaba-NLP/gte-multilingual-base> (дата обращения: 30.04.2025).
21. Rubert-tiny2 // Платформа Hugging Face.
URL: <https://huggingface.co/cointegrated/rubert-tiny2> (дата обращения: 30.04.2025).
22. Набор данных MMLU // Платформа Hugging Face.
URL: https://huggingface.co/datasets/alexandrainst/m_mmlu (дата обращения: 30.04.2025).
23. Wiki-ru-wordnet документация //
URL: <https://wiki-ru-wordnet.readthedocs.io/en/latest/> (дата обращения: 30.04.2025).

24. Библиотека для обработки естественного языка NLTK // URL: <https://www.nltk.org/> (дата обращения: 30.04.2025).

25. Liang C. et al. Distractor generation for multiple choice questions using learning to rank // Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications. 2018. Article number 28490.

USING MACHINE LEARNING TO ENHANCE TEST QUALITY

R. R. Miniukov¹ [0009-0008-1264-7107], M. M. Abramskiy² [0000-0003-3063-8948]

^{1, 2}Kazan (Volga Region) Federal University, Kazan, 420061, Russia

¹ramil.minyukov@gmail.com, ²mabramsk@kpfu.ru

Abstract

This study focuses on the application of machine learning methods to improve the quality of test items. The research includes a review of the subject area and the implementation of two enhancement methods: similar question retrieval and distractor quality assessment. The first method involves testing five transformer-based models for generating text embeddings and six clustering algorithms. The second method uses the same transformer models in combination with three classification algorithms. Experimental results demonstrated the high effectiveness of the proposed approaches in solving both tasks.

Keywords: *test item analysis, distractors, examination process, assessments, test quality improvement.*

REFERENCES

1. Chelyshkova M.B. Teoriya i praktika konstruirovaniya pedagogicheskikh testov. Moscow: Logos, 2002. 432 s.
 2. Hrich N., Azekri M., Khaldi M. An ai educational tool for detecting redundancy in distractors and items within multiple-choice tests // INTED2024 Proceedings. IATED, 2024. P. 6454–6458.
 3. Avanesov V.S. Teoriya i praktika pedagogicheskikh izmereniy. TsT i MKO UGTU-UI, 2005.
-

4. *Brusilovsky P., Miller P.* Web-based testing for distance education. In: P. DeBra and J. Leggett (Eds.) *Proceedings of WebNet'99, World Conference of the WWW and Internet*, Honolulu, HI, Oct. 24–30, 1999. AACE, P. 149–154.
5. *Tolstobrov A.P., Korzhik I.A.* Vozmozhnosti analiza i povysheniya kachestva testovykh zadaniy pri ispol'zovanii setevoy sistemy upravleniya obucheniem Moodle // *Vestnik VGU*. 2008. T. 2. S. 100–106.
6. *Alpatskaya E.V., Bubnov N.V., Minchenkov A.V.* Differentsiruyushchaya sposobnost' testovykh materialov dlya otsenki kachestva obucheniya // *Uchyonyye zapiski universiteta im. P. F. Lesgafta*. 2015. No. 11 (129). S. 9–14.
7. *Anastaszi A.* Psikhologicheskoe testirovanie. SPb.: Piter, 2007.
8. *Lord F.M.* Applications of item response theory to practical testing problems. Routledge, 2012.
9. *AlMahmoud R.H., Alian M.* The effect of clustering algorithms on question answering // *Expert Systems with Applications*. 2024. Vol. 243. Article number 122959.
10. *Zhang W.N. et al.* A topic clustering approach to finding similar questions from large question and answer archives // *PloS one*. 2014. Vol. 9, No. 3. Article number e71511.
11. *Alian M., Al-Naymat G.* Questions clustering using canopy-K-means and hierarchical-K-means clustering // *International Journal of Information Technology*, 2022. Vol. 14, No. 7. P. 3793–3802.
12. *Tarrant M., Ware J., Mohammed A. M.* An assessment of functioning and non-functioning distractors in multiple-choice questions: a descriptive analysis // *BMC medical education*. 2009. Vol. 9. P. 1–8.
13. *Moore S. et al.* An automatic question usability evaluation toolkit // *International Conference on Artificial Intelligence in Education*. Cham: Springer Nature Switzerland, 2024. P. 31–46.
14. First Quora Dataset Release: Question Pairs // Quora, 2017.
URL: <https://www.quora.com/q/quoradata/First-Quora-Dataset-Release-Question-Pairs> (last access: 30.04.2025).
15. Quora Question Pairs Russian // Kaggle, 2022.
URL: <https://www.kaggle.com/datasets/loopdigga/quora-question-pairs-russian> (last access: 30.04.2025).

16. Hugging Face – The AI Community [Электронный ресурс].
URL: <https://huggingface.co/> (last access: 30.04.2025).
17. All-MiniLM-L6-v2 // Hugging Face.
URL: <https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2> (last access: 30.04.2025).
18. Sentence-transformers/all-mpnet-base-v2 // Платформа Hugging Face
URL: <https://huggingface.co/sentence-transformers/all-mpnet-base-v2> (last access: 30.04.2025).
19. Paraphrase-multilingual-MiniLM-L12-v2 // Hugging Face.
URL: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2> (last access: 30.04.2025).
20. Gte-multilingual-base // Hugging Face.
URL: <https://huggingface.co/Alibaba-NLP/gte-multilingual-base> (last access: 30.04.2025).
21. Rubert-tiny2 // Платформа Hugging Face.
URL: <https://huggingface.co/cointegrated/rubert-tiny2> (last access: 30.04.2025).
22. Biblioteka dlya obrabotki estestvennogo yazyka
URL: <https://www.nltk.org/> (last access: 30.04.2025).
23. *Liang C. et al.* Distractor generation for multiple choice questions using learning to rank // Proceedings of the thirteenth workshop on innovative use of NLP for building educational applications, 2018. P. 284–290.

СВЕДЕНИЯ ОБ АВТОРАХ



МИНЮКОВ Рамиль Радикович – студент магистратуры Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Ramil Radikovich MINIUKOV – Master's student at the Institute of Information Technology and Intelligent Systems, Kazan Federal University.

email: ramil.minyukov@gmail.com

ORCID: 0009-0008-1264-7107



АБРАМСКИЙ Михаил Михайлович – директор Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета.

Mikhail Mikhailovich ABRAMSKIY – Director at the Institute of Information Technology and Intelligent Systems, Kazan Federal University.

email: mabramsk@kpfu.ru

ORCID: 0000-0003-3063-8948

Материал поступил в редакцию 5 мая 2025 года