

РЕЗУЛЬТАТЫ ДЕЯТЕЛЬНОСТИ ИССЛЕДОВАТЕЛЬСКИХ ЦЕНТРОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

УДК 004.8

О РАЗРАБОТКЕ ПРИКЛАДНЫХ РЕШЕНИЙ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ОБЕСПЕЧЕНИЯ ТЕХНОЛОГИЧЕСКОЙ БЕЗОПАСНОСТИ

© 2022 г. А. А. Масютин^{1,*}, А. В. Савченко¹, А. А. Наумов¹, С. В. Самсонов¹,
Д. Н. Тяпкин¹, Д. В. Беломестный¹, Д. С. Морозова¹, Д. А. Бадина¹

Представлено академиком РАН Г.И. Савиным

Поступило 28.10.2022 г.

После доработки 28.10.2022 г.

Принято к публикации 01.11.2022 г.

Основной миссией Исследовательского центра в сфере искусственного интеллекта НИУ ВШЭ (Центра ИИ) являются развитие и внедрение технологий искусственного интеллекта в различные сферы жизни человека и общества, отрасли науки и секторы экономики. В рамках деятельности Центра ИИ разрабатываются новые технологии искусственного интеллекта, позволяющие расширить область применения искусственного интеллекта; создаются программные инструменты и средства для применения искусственного интеллекта в отраслях науки и бизнеса, разрабатывается открытая программная библиотека методов искусственного интеллекта для решения задач, имеющих высокую социальную значимость.

Ключевые слова: технологии искусственного интеллекта, программные инструменты и средства для применения искусственного интеллекта, библиотека методов искусственного интеллекта

DOI: 10.31857/S2686954322070165

1. ВВЕДЕНИЕ

Деятельность Исследовательского центра в сфере искусственного интеллекта НИУ ВШЭ (Центра ИИ) реализуется в рамках федерального проекта “Искусственный интеллект”. Сроки реализации: 2021–2024 гг.

Главная задача Центра ИИ состоит в разработке прикладных решений на основе искусственного интеллекта для обеспечения технологической безопасности и технологического суверенитета Российской Федерации.

Основные направления исследований Центра ИИ: обработка естественного языка и интеллектуальные ассистенты; оптимизация краудсорсинговых платформ; ИИ в промышленности; рекомендательные системы; повышение качества аудио- и видеоданных; технологии для медиа; ИИ в образовании; ИИ в биоинформатике; ИИ и право. В общей сложности в Центре ИИ реализуется более 25 проектов, которые условно можно разделить на 2 большие группы: проекты для бизнеса и проекты для социально-значимой сферы

(рис. 1). Каждый проект соответствует мировым фронтам в конкретной технологической области.

Индустриальными партнерами Центра ИИ являются Сбербанк, Яндекс, MTS AI.

В качестве примеров проектов для бизнеса можно привести следующие:

Для Сбербанка:

- создание и оптимизация новых языковых моделей,
- разработка интеллектуальных систем с широкой областью применения: от задач повышения качества аудио до выявления признаков подозрительных транзакций.

Для Яндекса:

- алгоритмы оптимизации взаимодействия пользователей на различных платформах,
- решение, направленное на выявление субъективно неоднозначного (спорного) контента.

Для МТС:

- решения для телеком-рынка (например, оптимизация чат-ботов),
- применение генеративных моделей для восстановления качества изображений и аудио, а также возможности колоризации.

Для социально-значимой сферы в Центре разрабатываются решения для повышения качества

¹ Национальный исследовательский университет
“Высшая школа экономики”, Москва, Россия

*E-mail: amasyutin@hse.ru

ИИ для бизнеса		ИИ для социально-значимых задач
• Биоинформатика	• Банки, регулирование, экономика	• Медицина
• Повышение качества звука и изображения	• Экология и прогноз погоды	• Образование
• Технологии обработки естественного языка	• Технологии видеонаблюдения	• Медиапотребление и язык
• Генерация текстов	• Доставка данных по сети	• Рекомендательные сервисы
• Краудсорсинговые платформы	• Промышленная безопасность	• Право и этика

Рис. 1. Направления деятельности Центра ИИ.

образовательного процесса; модели машинного обучения для медицины, позволяющие снизить издержки при разработке медицинских препаратов; решения для судопроизводства и ответственного медиапотребления, а также ведутся исследования, направленные на решение вопросов в области этики ИИ.

Помимо трех индустриальных партнеров, ведется работа над привлечением дополнительных партнеров, которым интересны разработки Центра ИИ. Взаимодействие научных проектов и индустрии в дальнейшем приведет к появлению новых актуальных решений и продуктов на рынке.

Например, в 2022 г. подписаны четыре соглашения о сотрудничестве Центра ИИ с российскими компаниями, в том числе в сфере медицинских исследований, юридических услуг, туризма. Также ведутся переговоры с компаниями топливно-энергетического комплекса, с одним из крупнейших на российском рынке производителей лекарственных препаратов и другими компаниями.

За первое полугодие работы центра в этом году опубликован ряд статей, затрагивающих широкий спектр вопросов в области ИИ. Далее мы более подробно рассмотрим основные научно-прикладные результаты деятельности исследовательских проектов Центра ИИ.

2. ОБУЧЕНИЕ, ПОНИМАНИЕ И ОПТИМИЗАЦИЯ МОДЕЛЕЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Данный крупный проект “Обучение, понимание и оптимизация моделей искусственного интеллекта” имеет несколько важных направлений.

В рамках исследования по теме “Разработка и верификация алгоритмов и дополнительных математических моделей эффективного сэмплинга данных” был предложен эффективный подход к уменьшению дисперсии для аддитивных функционалов от цепей Маркова, основанный на новом представлении мартингала с дискретным временем. Исследуемый подход не требует знания стационарного распределения или конкретной структуры целевой плотности. Проведенный тщательный анализ теоретических свойств пред-

лагаемого алгоритм показал, что такой алгоритм позволяет добиться улучшения качества работы наивного Монте-Карло алгоритма при фиксированном вычислительном бюджете. Численная эффективность нового метода продемонстрирована для методов Монте-Карло по схеме марковской цепи (MCMC), основанных на динамике Ланжевена [1].

Также в рамках исследования цепей Маркова решалась проблема поиска оптимальной политики для марковских процессов принятия решений (MDP) с бесконечным горизонтом планирования. Для этой цели предлагается вариант стохастического зеркального спуска для задач выпуклого программирования с непрерывными Липшицевыми функционалами [2]. Важной деталью является возможность использовать неточные значения функциональных ограничений и вычислять значение двойственных переменных. Этот алгоритм был проанализирован в общем случае и получена оценка скорости сходимости, которая не накапливает ошибок во время работы метода.

Так был получен первый параллельный алгоритм для эргодичных MDP с усредненным вознаграждением, использующий генеративную модель, без сведения к дисконтированным MDP. Одной из ключевых особенностей представленного метода являются низкие коммуникационные затраты в контексте распределенного программирования в централизованной, даже очень большой, сети.

Ранее в других исследованиях было замечено, что обычный (суб)-градиентный метод может обрабатывать функциональные ограничения без дополнительных затрат с точки зрения суммарного количества итераций [3]. Способность к параллельным вычислениям имеет решающее значение для любого крупномасштабного приложения. В дальнейшем исследование развивалось в следующих направлениях:

- использование стохастических (суб)-градиентов;

- вычисление двойственных переменных для двойственной задачи Лагранжа (прямо-двойственность).

Первое направление имеет важное значение для масштабируемости высокопроизводитель-

ных приложений, поскольку вычисление точных градиентов может быть невозможным или долгим. Ценность второго типа улучшений сильно зависит от конкретного приложения, но всегда дает возможность использовать критерии останова, основанные на зазоре двойственности.

Благодаря результатам проекта, есть возможность объединить оба свойства и предложить стохастический субградиентный метод, который вычисляет двойные переменные без дополнительных затрат. Кроме того, в работе применяется предложенный прямо-двойственный стохастический зеркальный спуск к проблеме максимизации среднего вознаграждения в эргодичных марковских процессах принятия решений.

Следует отметить, что в Центре ИИ также ведутся исследования в области обучения с подкреплением, которые отражают актуальность общей деятельности Центра ИИ. В свою очередь, обучение с подкреплением (Reinforcement learning, RL) важно не только как объект исследований, но и как инструмент для решения важных практических задач. Обучение с подкреплением — один из видов машинного обучения. Ключевое отличие данного подхода от классического машинного обучения — это постоянное взаимодействие агента (алгоритма) со средой, от которой он получает обратную связь в виде поощрений и наказаний. Цель агента — максимизировать сумму наград, которые среда дает ему за “правильное” взаимодействие с ней. Для достижения этой цели агент должен балансировать между исследованием окружающей среды и использованием текущих знаний о ней для максимизации вознаграждений.

В рамках исследования RL [4, 5] был предложен алгоритм Bayes-UCBVI для обучения с подкреплением в табулярном эпизодическом MDP (марковском процессе принятия решений). Данный алгоритм является естественным обобщением алгоритма Bayes-UCBVI Кауфман и др., 2012 для многоруких бандитов. Данный метод использует апостериорную квантиль Q -функции как верхнюю доверительную границу для оптимальной Q -функции.

В данном проекте еще одно исследование было направлено на неасимптотический анализ алгоритмов линейной стохастической аппроксимации (LSA) с фиксированным размером шага. Это семейство методов возникает во многих задачах машинного обучения и используется для получения приближенных решений линейной системы $Ax = b$. В данной задаче величины A и b не наблюдаемы, однако доступна последовательность случайных наблюдений $\{(A_n, b_n)\}$, оценивающих A и b соответственно. Анализ был основан на новых результатах, касающихся моментов и оценок больших отклонений для произведений случайных

матриц, которые, как показано, являются точными. Данные оценки позволяют получить более точные оценки ошибки алгоритмов линейной стохастической аппроксимации при более слабых условиях на последовательность наблюдений $\{(A_n, b_n)\}$, чем в предыдущих работах. В частности, результаты не требуют симметричности случайных матриц $\{A_n\}$ [6].

Алгоритмы линейной стохастической аппроксимации имеют массу практических применений, в частности, в области обработки сигналов, шумоподавления, оптимизации, обучения с подкреплением. С помощью полученных результатов возможно получить новые оценки производительности RL-агентов в задаче оценивания политики, например, с помощью алгоритмов семейства TD (temporal differences).

3. НЕЙРОСЕТЕВЫЕ АЛГОРИТМЫ АНАЛИЗА ДИНАМИКИ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ И ВОВЛЕЧЕННОСТИ УЧЕНИКОВ НА ОСНОВЕ ДАННЫХ ВИДЕОНАБЛЮДЕНИЯ

В рамках проекта разрабатывается новый метод обработки видео для анализа эмоционального состояния участников видеоконференций, в том числе, учащихся в среде e-learning. На первом этапе применяются алгоритмы распознавания лиц, трекинга и кластеризации для извлечения последовательностей лиц каждого учащегося. Затем эмоциональные признаки лиц в каждом кадре извлекаются с помощью специальной вычислительно эффективной нейронной сети, которая предварительно обучена идентификации лиц и дообучена для классификации эмоций на фотографиях из набора AffectNet [7]. При этом в процессе обучения применяется специально разработанный робастный метод оптимизации. Показано, что полученные в результате векторы признаков лиц могут быть использованы для быстрого решения сразу нескольких задач: предсказания степени вовлеченности участников, их индивидуальных и групповых эмоций.

Разработанная нейросетевая модель может быть использована для обработки видео в режиме реального времени даже на мобильном устройстве пользователя без необходимости отправки видео их лиц на удаленный сервер или компьютер. Кроме того, продемонстрирована возможность подготовить краткое изложение хода онлайн-мероприятия с помощью создания коротких видеоклипов с фрагментами наиболее характерных эмоций каждого участника.

Экспериментальное исследование для наборов данных из конкурсов EmotiW показало, что предлагаемая сеть значительно превосходит существующие одиночные (не-ансамблевые) модели.

Основной вклад проекта состоит в следующем:

- ключевым компонентом предлагаемого подхода являются “легковесные” нейронные модели для распознавания выражений лиц на основе архитектур EfficientNet и MobileNet, которые могут использоваться для извлечения эмоциональных признаков лиц на фото и видео;

- предложена эффективная нейросетевая модель для одновременного обнаружения вовлеченности и распознавания эмоций на индивидуальном и групповом уровнях по видеоданным;

- представлена новая технология анализа лиц на видео в реальном времени.

Анализ вовлеченности и внимания может повысить эффективность и производительность проведения онлайн-встреч, а также обучения. Более того, была выявлена корреляция анализа выражения лиц и понимания.

Таким образом, целью проекта является разработка быстрой и точной методики для классификации эмоций и вовлеченности, которая может быть реализована в программном обеспечении на ноутбуках без мощных графических процессоров и или мобильных устройствах пользователей.

4. ДИАГНОСТИЧЕСКИЕ И АССИСТИВНЫЕ РЕЧЕВЫЕ ТЕХНОЛОГИИ НА ОСНОВЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Целью нейрохирургических вмешательств при опухолях головного мозга, рефрактерной эпилепсии, артериовенозных мальформаций является удаление патологической ткани при сохранении функционально необходимых областей, чтобы соответствующие функции, в том числе когнитивные [8], не были нарушены после нейрохирургии [9]. Одной из важнейших функций является обработка речи, которая лежит в основе человеческого общения, и ее нарушение негативно влияет на возвращение к работе, социальную интеграцию и общее качество жизни.

В рамках работы проекта “Диагностические и ассистивные речевые технологии на основе искусственного интеллекта” была исследована парадигма для предоперационного языкового картирования с использованием МРТ у русскоязычных людей и предоставлены методологические доказательства ее надежности при повторном тестировании.

Интраоперационное картирование языковой функции в головном мозге может быть дополнительно предоперационным картированием с помощью функциональной магнитно-резонансной томографии (ФМРТ/fMRI). Валидность парадигмы fMRI “языковой локализатор” в решающей степени зависит от выбора оптимальной языковой задачи и исходных условий. В этом исследовании представлен новый fMRI “языковой лока-

лизатор” на русском языке, использующий явное завершение предложения, задачу, которая все-сторонне задействует языковую функцию, включая как понимание, так и воспроизведение на уровне слов и предложений [10].

Парадигма была подтверждена на 18 неврологически здоровых добровольцах, которые участвовали в двух сеансах сканирования для оценки надежности повторного тестирования. На групповом уровне активизировались как передние, так и задние области, связанные с языком.

В целом исследование демонстрирует валидность и надежность задачи завершения предложения для отображения языковой функции в мозге. Кроме того, оно вносит вклад в общие данные о надежности fMRI при повторном тестировании.

5. РЕЗУЛЬТАТЫ ЦЕНТРА ИИ ЗА 2022 ГОД

Одним из ключевых показателей эффективности работы Центра ИИ является готовность некоторых проектов зарегистрировать результаты работ в виде программных обеспечений для ЭВМ и патентов на результаты интеллектуальной деятельности.

Так, например, в рамках проекта “Искусственный интеллект в макро моделировании и прогнозировании экономических процессов и финансовых взаимосвязей с учетом сентимента участников рынка” была разработана программа для анализа общерыночного сентимента инвесторов в трех тональностях по российскому рынку акций по сообщениям в каналах мессенджера, которая обеспечивает автоматизированный сбор, обработку и выявление сентимента (тональности) текстовых материалов инвестиционной тематики за указанный пользователем период времени при заданном минимальном периоде в один день.

Классификация материалов осуществляется с применением предобученной модели двунаправленной нейронной сети Google BERT, которая была усовершенствована с целью повышения качества классификации новостных текстов финансово экономического характера на наборе данных. Программа способна рассчитать 11 различных метрик общерыночного сентимента инвесторов по российскому рынку акций.

В данный момент продолжается работа по подготовке заявок еще несколькими проектами Центра ИИ в области компьютерного зрения, лингвистического анализа текстов и т.д.

6. ЗАКЛЮЧЕНИЕ

Результаты работы Центра ИИ подтверждают свою актуальность и позволяют решать различ-

ные задачи, стоящие перед российской экономикой:

— Языковые модели развиваются крайне динамично, в том числе и для русского языка. Для индустриальных партнеров в Центре ИИ ведутся работы по разработке моделей, оптимизирующих работу виртуальных ассистентов, чат-ботов. Результаты проектов позволят снизить рутинную нагрузку на службы поддержки клиентов, что в свою очередь положительно сказывается на снижении затрат на содержание колл-центров.

— В условиях ограничения доступа к технологическим рынкам повышается актуальность рекомендательных сервисов выявления технологических трендов и научных фронтиров. Именно поэтому в Центре ИИ имеется ряд проектов, которые занимаются разработкой рекомендательных систем.

— Так, в рамках Центра была разработана программная библиотека, рекомендуемая применение той или иной языковой модели, что позволяет при неизменном бюджете повышать качество решений NLP задач бизнеса, за счет выбора наиболее перспективных языковых моделей для экспериментов.

— Разработаны первые версии алгоритмов обработки открытых текстовых данных по компаниям, которые позволяют банковским организациям снижать операционные расходы на риск-анализ за счет нового потока данных о ESG-практиках компаний.

— Так как имеется ряд ограничений по доступу к образовательным продуктам, требуется создание уникальных импортозамещающих отечественных курсов. В 2022 г. было разработано 4 курса семинаров и лекций по тематике Центра ИИ, а также проведены школы по машинному обучению. К тому же одним из перспективных направлений Центра ИИ является использование искусственного интеллекта в образовании, в частности, разработка моделей персонализированного адаптивного обучения.

— Вследствие усложнения логистических цепочек, в том числе в сфере импортных лекарственных средств, повышается актуальность разработки отечественных лекарственных препаратов. В рамках Центра разрабатывается ПО для прогноза формы определенных участков антител. Решение позволит фарм-компаниям сократить число физических экспериментов, что позитивно скажется как на сроках разработки лекарств, так и на их стоимости.

Таким образом, результаты работы Центра должны лечь в основу создания и масштабирования прикладных решений в области искусственного интеллекта, которые находят применения в различных отраслях российской экономики.

СПИСОК ЛИТЕРАТУРЫ

1. *Belomestny D., Moulines E., Samsonov S.* Variance reduction for additive functional of Markov chains via martingale representations // *Statistics and Computing*. 2022. V. 32. № 1. Article 16. <https://doi.org/10.1007/s11222-021-10073-z>
2. *Tiapkin D., Gasnikov A.* Primal-Dual Stochastic Mirror Descent for MDPs // *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics (AISTATS 2022)*. 2022. V. 151. P. 9723–9740. <https://doi.org/10.48550/arXiv.2103.00299>
3. *Nemirovski A. and Yudin D.* Problem Complexity and Method Efficiency in Optimization // *A Wiley-Interscience publication*. Wiley. 1983.
4. *Tiapkin D., Belomestny D., Moulines É., Naumov A., Samsonov S., Tang Y., Valko M., Ménard P.* From Dirichlet to Rubin: Optimistic Exploration in RL without Bonuses, in *Proceedings of the 39th International Conference on Machine Learning*. 2022. V. 162. P. 21380–21431. <https://doi.org/10.48550/arXiv.2205.07704>
5. *Donald B. Rubin.* The bayesian bootstrap // *The annals of statistics*. 1981. P. 130–134.
6. *Durmus A., Moulines E., Naumov A., Samsonov S., Scaman K., Wai H.* Tight High Probability Bounds for Linear Stochastic Approximation with Fixed Stepsize // *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, pp. 1–12. <https://doi.org/10.48550/arXiv.2106.01257>
7. *Savchenko A., Savchenko L., Makarov I.* Classifying emotions and engagement in online learning based on a single facial expression recognition neural network// *IEEE Transactions on Affective Computing*, July 2022, pp. 1–12. <https://doi.org/10.1109/TAFFC.2022.3188390>
8. *Satoer D., Visch-Brink E., Dirven C., Vincent A.* Glioma surgery in eloquent areas: can we preserve cognition? *Acta Neurochi.* 2016. V. 158. P. 35–50. <https://doi.org/10.1007/s00701-015-2601-7>
9. *Duffau H.* The challenge to remove diffuse low-grade gliomas while preserving brain functions. *Acta Neurochir.* 2012. V. 154. P. 569–574. <https://doi.org/10.1007/s00701-012->
10. *Elin K., Malyutina S., Bronov O., Stupina E., Marinets A., Zhuravleva A., Dragoy O.* A New Functional Magnetic Resonance Imaging Localizer for Preoperative Language Mapping Using a Sentence Completion Task: Validity, Choice of Baseline Condition, and Test–Retest Reliability // *Frontiers in Human Neuroscience*. 2022. V. 16. P. 1–21. <https://doi.org/10.3389/fnhum.2022.791577>