

УДК 004.522:2

РАСКРЫТИЕ ПОТЕНЦИАЛА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ АНАЛИТИКИ БОЛЬШИХ ДАННЫХ



М.Б. Таджиева

*Преподаватель кафедры «Программное обеспечение информационных технологий»
Институт Телекоммуникаций и информатики Туркменистана
gurbangylyjovam@gmail.com*

М.Б. Таджиева

Окончила университет инженерных технологий Туркменистана имени Огуз хана. Область научных интересов связана с использованием и разработкой искусственного интеллекта, организацией учебного и научно-исследовательского процессов в институте.

Аннотация. В данной статье всесторонне рассматривается использование искусственного интеллекта (ИИ) в аналитике больших данных. Основное внимание уделяется методам машинного обучения и глубокого обучения, которые используются для разработки инновационных алгоритмов и решений в таких областях, как финансы, здравоохранение, окружающая среда и образование. В статье рассматриваются преимущества применения ИИ для анализа больших данных, такие как повышение эффективности и точности прогнозов, а также оптимизация решений. Однако в нем также подчеркиваются недостатки и проблемы, такие как обработка и безопасность информации, вопросы конфиденциальности и этические соображения. Возможности и технологические проблемы, связанные с обработкой огромных объемов данных, весьма многочисленны. В целом статья представляет собой сбалансированный взгляд на современный ИИ в условиях революции больших данных. Он информирует технических и нетехнических читателей о том, как добиться успеха на стыке больших данных и ИИ, реалистично оценивая проблемы, следуя принципам ответственного ИИ и уделяя особое внимание проектированию, ориентированному на человека.

Ключевые слова: Искусственного интеллекта (ИИ), большие данные, обучение, машинное обучение (МО).

Введение. Область анализа больших данных с помощью искусственного интеллекта охватывает различные методы, в основном относящиеся к областям машинного обучения и глубокого обучения. Эти методы играют ключевую роль в современных приложениях ИИ для анализа огромных наборов данных, предлагая больше, чем просто инструменты — они представляют собой основу анализа данных в эпоху ИИ. Машинное обучение — важная область искусственного интеллекта — использует статистические методы, позволяющие компьютерам выполнять определенные задачи на основе предоставленных данных. Этот подход автоматизирует построение аналитических моделей, основанных на принципе, согласно которому системы могут обучаться на основе данных, выявлять закономерности и принимать решения с минимальным вмешательством человека. Область машинного обучения можно условно разделить на три типа: контролируемое обучение, неконтролируемое обучение и обучение с подкреплением.

При контролируемом обучении алгоритмы обучаются на маркированных данных, которые включают в себя входные данные и соответствующие им правильные выходные данные. Алгоритм учится сопоставлять эти входные данные с выходными данными, что широко используется в приложениях, где исторические данные предсказывают вероятные будущие события, например, при обнаружении мошенничества в банковских транзакциях. С другой стороны, неконтролируемое обучение имеет дело с немаркированными данными.

Система пытается изучить структуру и закономерности таких данных, что обычно используется в исследовательском анализе данных для поиска скрытых закономерностей или группировок в данных, например, при сегментации клиентов. Обучение с подкреплением подразумевает, что агент учится вести себя в окружающей среде, выполняя действия и наблюдая за результатами. Этот метод в основном применяется в сценариях, где принятие решений является последовательным, а цель — долгосрочной, например, в стратегиях робототехники.

Глубокое обучение — специализированный подвид машинного обучения — основано на искусственных нейронных сетях с несколькими слоями, что позволяет модели обучаться и принимать разумные решения самостоятельно. Он произвел революцию в области искусственного интеллекта благодаря своей способности обрабатывать и анализировать сложные и неструктурированные данные, такие как изображения, звук и текст.

Сверточные нейронные сети в основном используются для обработки изображений, видео, аудио и других двумерных данных, прекрасно справляясь с такими задачами, как распознавание изображений и видео, классификация изображений и анализ медицинских изображений. Рекуррентные нейронные сети, используемые для последовательных данных, таких как анализ временных рядов, языковое моделирование и перевод текста, отличаются от традиционных нейронных сетей тем, что имеют память для хранения предыдущих входных данных во внутренней памяти, что делает их идеальными для последовательных данных. Трансформеры, изначально разработанные для задач обработки естественного языка, таких как перевод и реферирование текстов, произвели революцию в машинном обучении. Они известны своей эффективностью в распараллеливании обработки данных и обработке долгосрочных зависимостей в данных.

Интеграция машинного обучения и глубокого обучения в анализ больших данных открыла новые горизонты в исследованиях и практическом применении искусственного интеллекта. Эти методы меняют наш подход к большим наборам данных и их понимание — от предиктивной аналитики в финансах до расширенного распознавания изображений в здравоохранении. По мере развития технологий увеличивается потенциал более сложного и эффективного анализа больших данных на основе искусственного интеллекта, что обещает дальнейшие инновации и прорывы в различных областях.

Применение искусственного интеллекта для анализа больших данных в различных отраслях.

Использование искусственного интеллекта для анализа больших данных получило широкое распространение во многих отраслях. У каждого сектора есть свои уникальные проблемы и требования, и ИИ предлагает индивидуальные решения для эффективного удовлетворения этих потребностей. В финансовом секторе ИИ играет решающую роль в анализе больших объемов данных для различных целей. Он играет важную роль в прогнозировании цен на акции, оценке кредитных рисков и выявлении мошеннических действий. Методы машинного обучения, включая классификацию и регрессию, применяются для анализа финансовых данных, помогая принимать стратегические решения. Эти анализы на основе искусственного интеллекта предоставляют информацию, имеющую решающее значение для управления рисками, инвестиционных стратегий и обслуживания клиентов в финансовой отрасли.

Здравоохранение — еще одна область, где влияние ИИ на анализ больших данных имеет преобразующее значение. Алгоритмы ИИ способны выявлять закономерности в медицинских данных, которые могут помочь в ранней диагностике заболеваний, разработке новых лекарств и определении наиболее эффективных методов лечения. Глубокое обучение, в частности, применяется для анализа медицинских изображений и генетических данных, предлагая расширенные возможности диагностики и персонализированной медицины. Это не только улучшает результаты лечения пациентов, но и способствует развитию медицинских исследований. Роль ИИ в анализе экологических данных

становится все более важной в борьбе с изменением климата и сохранением окружающей среды. Методы машинного обучения используются для прогнозирования погодных условий, мониторинга состояния экосистем и разработки стратегий по сокращению выбросов загрязняющих веществ. Анализируя данные, связанные с климатом, загрязнением и биоразнообразием, ИИ помогает разрабатывать более эффективную экологическую политику и стратегии сохранения природы.

Сфера образования переживает стремительные перемены благодаря интеграции искусственного интеллекта. Алгоритмы ИИ анализируют индивидуальные характеристики учащихся, включая их знания, навыки и стили обучения, чтобы предоставлять персонализированные учебные материалы и задания. Такой адаптивный подход к обучению помогает учащимся учиться в своем собственном темпе и сосредотачиваться на тех областях, где им необходимо улучшить знания. Автоматизированные системы оценивания, работающие на основе искусственного интеллекта, эффективно оценивают ответы учащихся, особенно при решении задач с множественным выбором или математических задач, экономя преподавателям драгоценное время. Алгоритмы ИИ также предлагают персонализированные рекомендации по курсам, видео лекциям, учебникам и другим образовательным ресурсам на основе предпочтений и интересов учащихся. Для обеспечения академической честности ИИ используется для обнаружения плагиата в эссе и научных работах, а также для мониторинга студентов во время онлайн-экзаменов с целью предотвращения списывания. Виртуальные помощники и чат-боты, работающие на основе искусственного интеллекта, отвечают на вопросы студентов, помогают с домашними заданиями и предоставляют информацию о курсе. Это особенно полезно в массовых открытых онлайн-курсах, где личное взаимодействие с преподавателями может быть ограничено. Возможности предиктивной аналитики ИИ распространяются на прогнозирование успеваемости учащихся, оптимальной нагрузки и рисков отсева путем анализа таких данных, как посещаемость, оценки и активность в системах управления обучением.

Технологии искусственного интеллекта также используются для развития социальных навыков и эмоционального интеллекта в виртуальных средах, что особенно полезно для учащихся с особенностями развития или социальными тревожностями. Более того, само по себе образование в области ИИ является развивающейся областью,

совместно с образовательными учреждениями разрабатывают курсы и программы, направленные на подготовку студентов к карьере в области искусственного интеллекта, машинного обучения и робототехники. Наконец, технологии искусственного интеллекта анализируют данные учащихся, такие как успеваемость, интересы и стили обучения, чтобы предоставить ценную обратную связь педагогам. Это помогает учителям совершенствовать методы обучения и адаптировать материалы для максимальной эффективности. ИИ также используется при создании контента, такого как видеоролики, анимация, тексты и визуализации, что позволяет педагогам сосредоточиться на стратегиях обучения, а не на рутинных задачах.

Поскольку искусственный интеллект в образовании продолжает развиваться, он привносит инновации в традиционные методы обучения, предлагая возможности для персонализированного обучения, повышения эффективности преподавания и улучшения взаимодействия между преподавателями и учениками. Однако при разработке и внедрении решений на основе ИИ в образовании важно учитывать потенциальные проблемы, такие как конфиденциальность данных, этические вопросы и неравный доступ к технологиям.

Аппаратные и программные технологии для применения искусственного интеллекта в анализе больших данных.

В сфере искусственного интеллекта и анализа больших данных синергия между передовыми аппаратными и программными технологиями играет ключевую роль. Эти технологии не только обеспечивают необходимую вычислительную мощность, но и

предоставляют необходимые инструменты и структуры для эффективной обработки и анализа данных с помощью ИИ. Высокопроизводительные вычислительные системы находятся на переднем крае аппаратных технологий, позволяющих применять ИИ в работе с большими данными. Эти системы, включая суперкомпьютеры и распределенные вычислительные среды, обладают способностью обрабатывать и анализировать большие наборы данных гораздо быстрее, чем стандартные вычислительные системы. Графические процессоры, изначально разработанные для рендеринга графики, стали краеугольным камнем в обработке ИИ, особенно для алгоритмов глубокого обучения. Их возможности параллельной обработки особенно подходят для выполнения сложных вычислений, необходимых в ИИ. Тензорные процессоры, специально разработанные для задач машинного обучения, оптимизируют операции ИИ, особенно при вычислениях нейронных сетей. Кроме того, развивающаяся область квантовых вычислений открывает будущее, в котором сложные вычисления ИИ можно будет выполнять за доли текущего времени, что обещает значительный скачок в возможностях обработки данных. С точки зрения программного обеспечения незаменимыми стали такие фреймворки машинного обучения, как TensorFlow, PyTorch и Keras. Они предлагают готовые функции и структуры, которые упрощают создание и обучение моделей машинного обучения. Инструменты обработки больших данных, такие как Apache Hadoop и Apache Spark, обеспечивают необходимую структуру для распределенного хранения и обработки, что имеет решающее значение для обработки и анализа огромных объемов данных. Инструменты визуализации данных, такие как Tableau и Power BI, играют важную роль в переводе результатов анализа с помощью ИИ в понятные и применимые на практике форматы. Они позволяют преобразовывать сложные шаблоны данных в более понятные визуальные представления.

Платформы облачных вычислений, такие как AWS, Google Cloud и Microsoft Azure, играют решающую роль, предлагая масштабируемые и гибкие ресурсы. Эти платформы предоставляют целый ряд услуг — от хранения данных до расширенных возможностей машинного обучения, позволяя пользователям получать доступ к вычислительной мощности высокого уровня без необходимости в обширной физической инфраструктуре. Эффективное применение ИИ в анализе больших данных зависит от бесшовной интеграции этих аппаратных и программных технологий. Такая интеграция не только обеспечивает эффективную обработку больших наборов данных, но и позволяет применять сложные алгоритмы и получать полезные выводы. Это требует стратегического подхода для обеспечения совместимости и оптимальной производительности различных технологий.

Ландшафт аппаратных и программных технологий для ИИ в анализе больших данных разнообразен и динамичен. Поскольку эти технологии продолжают развиваться, они будут способствовать дальнейшему расширению возможностей и сфер применения ИИ в различных областях, делая анализ больших данных более эффективным, точным и информативным. Понимание и эффективное использование этих технологий является ключом к раскрытию полного потенциала ИИ в сфере больших данных. На схеме, представленной на рисунке, показано взаимодействие между ключевыми компонентами, участвующими в применении ИИ и машинного обучения для анализа больших данных. Все начинается с того, что пользователь генерирует новые данные, которые собираются различными источниками данных. Источники данных создают новые объекты для сохранения необработанных данных в компоненте хранения данных.

Параллельно компонент модели обработки данных и машинного обучения извлекает данные из хранилища для предварительной обработки и очистки. Затем он инициализирует модель машинного обучения, обучает ее на предварительно обработанных данных, настраивает гиперпараметры и завершает модель. После обучения модели ее разворачивают путем создания обслуживающего компонента модели для составления прогнозов на основе новых данных.

Параллельно с этим выполняется анализ новых данных путем получения новых данных из хранилища, составления прогнозов с использованием развернутой модели, визуализации результатов и предоставления аналитических сведений пользователю.

Компонент развертывания модели также получает данные в реальном времени из источников, делает прогнозы и отправляет их на визуализацию.

В целом, он дает общее представление о применении ИИ и машинного обучения к большим данным, разбивая этапы и демонстрируя взаимодействие между компонентами.

Заключение. Анализ больших данных с использованием искусственного интеллекта и машинного обучения открыл новые горизонты для извлечения полезной информации из больших и разнообразных наборов данных во многих отраслях и приложениях. Однако, каким бы многообещающим ни казалось это сочетание, существуют определенные ограничения и риски, которые необходимо признать и смягчить для его эффективного и этичного использования.

Одной из ключевых задач является разработка моделей ИИ/МО (машинное обучение), которые могут предоставлять объяснимые прогнозы и решения вместо непостижимых «черных ящиков». Такие методы, как диагностические объяснения с использованием локально интерпретируемых моделей, могут пролить свет на то, почему модели делают определенные прогнозы, повышая доверие и подотчетность. Кроме того, необходимо использовать такие методы, как состязательное обучение, проверки надежности и анализ чувствительности, чтобы уменьшить предвзятость и построить более обобщенные модели, менее подверженные неожиданным ошибкам.

Список литературы

- [1] Z.M.Gizatullin, M.S.Shkinderov, R.R.Mubarakov, Proceedings of the 2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering, 1350-1353, 2022.
- [2] R. F. Gibadullin, N. S. Marushkai, 2021 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 404-409, 2021.
- [3] R. F. Gibadullin, I. S. Vershinin, R. Sh. Minyazev, 2017 International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM), 1-6, 2017.

UNLOCKING THE POTENTIAL OF ARTIFICIAL INTELLIGENCE FOR BIG DATA ANALYTICS

M.Tajieva

*Lecturer of the Department of Software of information technologies,
The Institute of Telecommunications and Informatics of Turkmenistan*

Annotation. This article comprehensively examines the use of artificial intelligence (AI) in big data analytics. It focuses on machine learning and deep learning methods that are leveraged to develop innovative algorithms and solutions across domains like finance, healthcare, environment, and education. The article discusses the benefits of applying AI to big data analysis such as improved efficiency and accuracy of predictions, as well as optimization of decisions. However, it also highlights downsides and challenges such as information processing and security, privacy concerns, and ethical considerations.

Keywords: Artificial intelligence (AI), big data, learning, machine learning (ML).