

Секция «Искусственный интеллект и цифровая трансформация в бизнесе и государственном управлении»

Применение ИИ для динамической классификации документов в системах управления контентом

Научный руководитель – Кочкаров Расул Ахматович

Корчагин Сергей Алексеевич

Кандидат наук

Финансовый университет, Факультет информационных технологий и анализа больших данных, Москва, Россия

E-mail: korchaginser@gmail.com

В последние годы развитие технологий искусственного интеллекта (ИИ) существенно изменило многие сферы государственного сектора и бизнеса, включая управление контентом [1]. Одним из наиболее перспективных направлений является динамическая классификация документов. Эта технология позволяет автоматизировать процесс обработки информации, облегчая доступ к данным и повышая эффективность работы организаций. Динамическая классификация документов с использованием ИИ предлагает возможность систематического анализа больших объемов информации, что особенно актуально в условиях быстрого увеличения объема данных. Традиционные методы классификации часто требуют значительных затрат времени и ресурсов, в то время как системы, оснащенные ИИ, могут обрабатывать документы в реальном времени, обеспечивая более высокую скорость и точность. Основной принцип работы динамической классификации основан на алгоритмах машинного обучения, которые способны выявлять паттерны и взаимосвязи в данных [3]. С помощью обучения на больших выборках документов, ИИ может самостоятельно определять категории и подкатегории, в которые должен быть отнесен каждый документ. Это не только облегчает процесс организации информации, но и минимизирует вероятность ошибок, часто возникающих при ручной классификации [4]. Применение ИИ в системах управления контентом также позволяет адаптироваться к изменяющимся требованиям бизнеса.

В работе рассмотрены современные технологии ИИ для улучшения классификации и управления документами в системах управления контентом (CMS). Проведен анализ использования систем автоматического распознавания документов для оптимизации процессной работы в бизнесе и государственных учреждениях. Исследованы популярные алгоритмы глубокого обучения, для точной идентификации и классификации документов по заранее определённым категориям. Проведено сравнение различных архитектур нейронных сетей. Для решения задачи динамической адаптации процесса классификации в зависимости от контекста и содержания загружаемых материалов использованы методы обработки естественного языка (NLP), что позволило повысить эффективность анализа текстового содержания документов. Рассмотрены практические кейсы успешного применения ИИ для динамической классификации документов в различных отраслях. Приводятся следующие примеры: финансовый сектор (обработка заявок и договоров), государственный сектор (организация работы с большими объемами документов, путём автоматического распределения по категориям в электронных архивах). Предложены вопросы для обсуждения преимуществ, которые дают технологии ИИ в части автоматической обработки документов, такие как сокращение времени на обработку документов, снижение числа ошибок, повышение уровня безопасности, улучшение доступа к информации.

В исследовании показано, что применение ИИ для динамической классификации документов в системах управления контентом становится одним из ключевых аспектов эф-

фективного управления информацией. Приводятся примеры применения систем управления контентом (CMS): обработка, хранения и извлечение нужной информации. Проведены расчеты временных затрат и человеческих ресурсов по сравнению с традиционными методами классификации. Рассмотрены подходы искусственного интеллекта, в частности технологии машинного обучения и обработки естественного языка (NLP) для динамической классификации документов. Установлено, что автоматизированная классификация открывает возможность использования моделей машинного обучения, обученных на исторических данных, автоматически классифицировать новые документы в режиме реального времени. В рамках исследования рассмотрены такие алгоритмы, как наивный байесовский классификатор, деревья решений, методы опорных векторов, нейронные сети и ансамблевые модели. С использованием данных алгоритмов был проведен анализ текстовых документов, извлечены сущности и найдены семантические связи. Также были рассмотрены алгоритмы обучения с подкреплением, которые использовались для динамической настройки моделей классификации в зависимости от изменений в типах документов или требований пользователей. Такой подход позволил системе адаптироваться и поддерживать высокую степень точности.

При анализе алгоритмов машинного обучения для динамической классификации документов в системах управления контентом используются несколько ключевых метрик, которые помогают оценить эффективность и качество классификации. Основной метрикой является точность (Accuracy), которая рассчитывается как отношение правильно классифицированных документов ко всем документам в тестовом наборе, предоставляя общее представление о том, насколько хорошо модель работает. Полнота (Recall) – это другая важная метрика, показывающая, какую долю от всех положительных классов модель смогла правильно классифицировать. Она особенно важна в случаях, когда критично обнаруживать все случаи позитивного класса. Точность (Precision) указывает процент правильно предсказанных положительных классов из всех предсказанных положительных, и она особенно полезна, когда важно избегать ложных срабатываний. F1-меры представляют собой гармоническое среднее между полнотой и точностью, что позволяет получить единую метрику, балансируя эти два показателя, и это особенно полезно в случае несбалансированных классов. Использование этих метрик позволяет всесторонне оценить качество и эффективность алгоритмов классификации документов, выбрать наилучший подход для конкретной задачи и выявить потенциальные области для улучшения.

Отдельно рассмотрены современные модели ИИ, которые могут быть настроены на непрерывное обучение, что позволяет таким моделям оставаться актуальными с новыми данными, адаптируясь и улучшая имеющиеся алгоритмы.

В рамках исследования рассмотрены кейсы отечественных и международных компаний по внедрению технологий ИИ для динамической классификации документов, в том числе в банковской сфере, где используются для классификации финансовых документов и кредитных заявок, в здравоохранении для управления пациентскими записями и в юридической сфере для автоматизации анализа документов. В результате проведенного анализа установлен ряд преимуществ, которые можно получить используя ИИ для динамической классификации документов: скорость и эффективность обработки, высокую степень точности, снижение ошибок, улучшение клиентского опыта и пр. Анализ кейсов показал, что автоматизированные системы обработки документов позволяют значительно сократить время, потраченное на классификацию документов, а высокоточные модели ИИ снижают вероятность ошибок, связанных с человеческим фактором. В результате, внедрение технологий ИИ в процессы классификации становится важной частью успешной стратегии цифровой трансформации для организаций, стремящихся повысить общую эффективность и качество своих услуг.

Стоит отметить, что при внедрении ИИ для динамической классификации документов возникают определённые сложности. Следует обратить внимание на потенциальные проблемы, связанные с этическими аспектами, защитой данных и необходимостью обучения персонала для работы с новыми технологиями. В исследовании приводятся рекомендации по внедрению ИИ в существующие системы управления контентом, включая создание инфраструктуры для обработки больших данных, выбор подходящих алгоритмов и необходимость постоянного обучения моделей для повышения их точности и эффективности. Искусственный интеллект открывает новые горизонты в системах управления контентом и инвестиции в развитие ИИ-технологий могут значительно улучшить управление информацией, что в конечном итоге приведет к повышению конкурентоспособности организаций в различных сферах.

Применение ИИ для динамической классификации документов представляет собой мощный инструмент, который способствует повышению эффективности работы с контентом, улучшает взаимодействие с пользователями и экономит ресурсы. Все эти аспекты подчеркивают важность интеграции технологий ИИ в системы управления контентом в современном цифровом мире.

Источники и литература

- 1) Апатова Н. В. Управление процессами цифровой трансформации бизнеса // Ученые записки Крымского федерального университета имени В.И. Вернадского. Экономика и управление. – 2022. – Т. 8. – №. 2. – С. 3-8.
- 2) Митрофанова О. А., Атугодаге М. М. Динамическое тематическое моделирование русскоязычного корпуса юридических документов // Terra Linguistica. – 2023. – Т. 14. – №. 1. – С. 70-87.
- 3) Tachinina O., Lysenko A., Kutieпов V. Classification of modern unmanned aerial vehicles // Electronics and Control Systems. – 2022. – Т. 4. – №. 74. – С. 79-86.
- 4) Alothman A. F., Wahab Sait A. R. Managing and Retrieving Bilingual Documents Using Artificial Intelligence-Based Ontological Framework // Computational Intelligence and Neuroscience. – 2022. – Т. 2022. – №. 1. – С. 4636931.