

быть не всегда точной при выставлении оценок обучающихся, так как искусственный интеллект не обладает человеческим фактором, а значит, не имеет эмпатии, критического мышления, социальных навыков и ограничен в творчестве. В-третьих, остро возникает вопрос этики: неравномерный доступ к технологиям искусственного интеллекта разных групп учеников и обеспечение конфиденциальности личных данных обучающихся.

На основе полученных результатов можно сделать вывод, что генеративный искусственный интеллект действительно является сильным инструментом в сфере образования, но не может быть единственным и самостоятельным, так как вклад человека по-прежнему играет важную роль в подготовке будущего поколения, развитии у них критического мышления и творчества, самостоятельности. Технология машинного обучения требует разработки нормативных документов, регулирующих этические стандарты, безопасность и справедливость. Необходимо гармонично сочетать возможности ГИИ с человеческими, так как он является дополняющим инструментом, а не заменяющим.

Смирнов В.Н.

кандидат педагогических наук
Ленинградский областной филиал
Санкт-Петербургского университета МВД России (п. Мурино)

Применение технологий обработки естественного языка для анализа документов и показаний в органах внутренних дел

В эпоху цифровой трансформации и стремительного роста объемов текстовой информации органы внутренних дел сталкиваются с необходимостью обрабатывать и анализировать огромные массивы документов – протоколов, показаний, отчетов, переписки и др. Традиционные «ручные» методы работы с такими неструктурированными данными становятся все более трудозатратными и неэффективными. В этих условиях на помощь приходят технологии обработки естественного языка (Natural Language Processing, далее – NLP), позволяющие автоматизировать многие рутинные процессы и высвободить человеческие ресурсы для решения нетривиальных задач.

NLP представляет собой одно из наиболее динамично развивающихся направлений искусственного интеллекта, нацеленное на анализ, понимание и генерацию человеческих языков¹. Методы NLP находят все более широкое применение в таких сферах, как здравоохранение, финансы, маркетинг, образование и др. Однако потенциал этих технологий для повышения

¹ Raaijmakers S. Deep Learning for Natural Language Processing. Manning, 2022. 296 p.

эффективности работы правоохранительных органов пока еще недооценен и требует более глубокого изучения.

Обработка естественного языка охватывает широкий спектр методов и алгоритмов, нацеленных на анализ текстов на разных уровнях – от отдельных символов и слов до семантики целых предложений и документов. Рассмотрим некоторые базовые NLP-подходы, которые могут быть полезны для автоматизации анализа текстовой информации в органах внутренних дел.

Токенизация (Tokenization) – это процесс разбиения текста на минимальные единицы (токены) – слова, знаки препинания, числа и т.д. Это необходимый этап предобработки текста перед применением других NLP-алгоритмов¹. При кажущейся простоте токенизация может быть нетривиальной задачей для некоторых языков и предметных областей (например, при наличии большого количества аббревиатур, имен собственных, дефисных написаний и т.п.).

Удаление стоп-слов (Stop words removal) предполагает фильтрацию из текста неинформативных и часто встречающихся слов (предлоги, союзы, местоимения и т.д.). Это позволяет уменьшить размерность анализируемых данных и сфокусироваться на действительно важных элементах. Однако следует учитывать, что в некоторых случаях (например, при анализе тональности текста) удаление стоп-слов может исказить смысл.

Стемминг (Stemming) и лемматизация (Lemmatization) направлены на приведение различных словоформ (падежей, склонений, спряжений) к начальной форме слова. Стемминг реализуется путем «отсечения» окончаний по формальным правилам без учета контекста и часто дает некорректный результат. Лемматизация использует более сложные лингвистические алгоритмы и словари, позволяя получать осмысленные начальные формы.

Векторное представление текста (Vector Space Model) – это способ преобразования текста в числовой вектор фиксированной размерности. Самый простой подход – «мешок слов» (Bag of Words) – представляет документ в виде вектора частот входящих в него слов. Более продвинутые методы (TF-IDF, word2vec и др.) учитывают значимость и семантику слов. Векторизация текста необходима для применения алгоритмов машинного обучения.

Тематическое моделирование (Topic Modeling) позволяет выявлять скрытые темы и кластеризовать документы по их содержанию. Наиболее популярным методом является латентное размещение Дирихле (LDA), которое в неконтролируемом режиме находит оптимальное распределение слов по темам и документов по темам. Это дает возможность быстро сориентироваться в больших коллекциях текстов.

¹ Ваш путеводитель по миру NLP (обработке естественного языка) / Habr. URL: <https://habr.com/ru/companies/otus/articles/705482/> (дата обращения: 14.10.2024).

Помимо перечисленных базовых подходов, современные NLP-системы используют более сложные модели глубокого обучения, позволяющие учитывать контекст и извлекать глубинную семантику текста. Однако даже простые методы могут дать ощутимый выигрыш по сравнению с «ручной» обработкой документов.

Для того чтобы оценить потенциал NLP-технологий для органов внутренних дел, полезно рассмотреть примеры их успешного использования в других сферах деятельности. Это позволит понять спектр возможных приложений и оценить достигаемые эффекты.

Одной из наиболее перспективных отраслей для NLP является здравоохранение. Методы обработки естественного языка здесь применяются для автоматического извлечения информации из медицинских записей, построения диагностических рекомендательных систем, анализа отзывов пациентов и т.д. Например, сервис Amazon Comprehend Medical использует NLP для выявления упоминаний болезней, лекарств и лечебных процедур в текстовой медицинской документации.

Другая область активного внедрения NLP – это финансы. Алгоритмы анализа тональности текстов (sentiment analysis) позволяют по новостям, публикациям в соцсетях и другим источникам оценивать настроения рынка и делать прогнозы. NLP-модели используются для автоматической обработки запросов клиентов, выявления случаев мошенничества, анализа кредитных историй и др. Так, американский стартап Amenity Analytics применяет NLP для извлечения инвестиционных идей из потоков новостей и финансовых отчетов.

В сфере управления персоналом NLP-системы берут на себя задачи по автоматизации подбора и оценки кандидатов. Анализируя резюме и профили соискателей, рекрутеры могут быстро отфильтровывать подходящих специалистов по ключевым навыкам и компетенциям. Чат-боты на основе NLP способны давать ответы на типовые вопросы кандидатов и даже проводить первичные интервью¹.

Очевидно, что многие задачи, с которыми сталкиваются сотрудники органов внутренних дел, связаны с анализом больших объемов неструктурированной текстовой информации. Это могут быть свидетельские показания, протоколы осмотров мест происшествий, материалы уголовных дел, оперативные сводки, отчеты, нормативные документы и т.д. Обработка всех этих источников вручную крайне трудозатратна и нередко приводит к ошибкам и упущениям.

¹ Технология обработки естественного языка / Microsoft Learn. URL: <https://learn.microsoft.com/ru-ru/azure/architecture/data-guide/technology-choices/natural-language-processing> (дата обращения: 14.10.2024).

Представляется, что внедрение NLP-технологий способно значительно повысить эффективность работы с текстовыми данными в правоохранительной сфере. Приведем несколько гипотетических сценариев использования методов обработки естественного языка.

1. Автоматическое реферирование и поиск ключевой информации в больших массивах документов (материалах дел, архивах и др.). NLP-алгоритмы могут выделять главные факты, действующих лиц, даты, суммы и т.п., существенно экономя время следователей и оперативных сотрудников.

2. Выявление неявных взаимосвязей между делами, фигурантами и событиями путем кросс-документного анализа. Сопоставляя информацию из разных источников, NLP-система может находить неочевидные ассоциации и генерировать новые зацепки для расследования.

3. Автоматическая категоризация обращений граждан и других входящих документов. NLP-классификаторы способны маршрутизировать запросы в нужные подразделения, выявлять срочные сигналы, требующие оперативного реагирования, а также отсеивать спам и нерелевантные обращения.

4. Выявление противоречий и оценка достоверности информации в свидетельских показаниях. Сравнивая показания разных лиц с помощью NLP-алгоритмов, можно находить несоответствия в деталях, оценивать психоэмоциональное состояние говорящих, распознавать попытки введения в заблуждение.

5. Составление психолингвистического портрета по образцам речи и текстов. На основе специфических лексических, синтаксических и стилистических паттернов NLP-модели потенциально способны делать выводы о поле, возрасте, образовании, психотипе и других характеристиках автора текста.

Безусловно, внедрение NLP в работу органов внутренних дел сопряжено с рядом технических, организационных и этических вызовов. Необходимо обеспечить высокое качество моделей, их адаптацию к специфической предметной области, соблюдение требований информационной безопасности и приватности данных. Важно выстроить эффективное взаимодействие между техническими специалистами и конечными пользователями – следователями, дознавателями, оперативными сотрудниками.

Тем не менее мы убеждены, что преимущества от использования NLP в правоохранительной системе перевешивают возможные риски и издержки. Алгоритмы обработки естественного языка способны взять на себя значительную часть рутинной аналитической работы, высвобождая человеческие ресурсы для принятия ответственных решений. При этом NLP следует рассматривать не как замену, а как интеллектуальный инструмент, расширяющий возможности сотрудников.

Подводя итог, можно сказать, что технологии обработки естественного языка открывают новые возможности для повышения эффективности работы

органов внутренних дел в части анализа текстовой информации. Методы NLP позволяют автоматизировать многие трудоемкие процессы, такие как реферирование документов, выявление значимых фактов, категоризация обращений, поиск неявных взаимосвязей и т.д.

Дальнейшие исследования в этой области должны быть направлены на создание специализированных NLP-моделей и приложений с учетом отраслевой специфики и нормативно-правовых требований. Необходимо организовать тесное взаимодействие между разработчиками и практическими работниками для выработки оптимальных сценариев использования NLP в реальной оперативно-служебной деятельности. Только так можно будет в полной мере раскрыть потенциал обработки естественного языка как интеллектуального помощника в работе органов внутренних дел.

Спичкин М.Ю.

Воронежский институт ФСИН России

Некоторые особенности противодействия преступлениям в сфере незаконного оборота наркотических средств, совершаемым посредством сети Интернет

В соответствии с Конституцией РФ жизнь и здоровье человека являются приоритетной основой процветания российского общества (ст. 2 «Человек, его права и свободы являются высшей ценностью. Признание, соблюдение и защита прав и свобод человека и гражданина – обязанность государства»). Как отмечается в Указе Президента РФ «О стратегии национальной безопасности РФ», к основным источникам угроз национальной безопасности относится «деятельность транснациональных преступных организаций и группировок, связанная с незаконным оборотом наркотических средств и психотропных веществ»¹. Следовательно, борьба с незаконным оборотом наркотиков относится к важнейшим направлениям деятельности государства.

Доступность наркотиков максимально проста. Каждый человек, способный использовать гаджеты (ПК, ноутбук, планшет, смартфон и т.д.), с легкостью может приобрести всевозможные разновидности наркотических средств или психотропных веществ. При этом обладать какими-либо специальными навыками и связями в преступной среде вовсе необязательно. Особенно опасно для общества, что под ударом находятся несовершеннолетние,

¹ Стратегия национальной безопасности Российской Федерации : Указ Президента Российской Федерации от 31.12.2015 № 683.