

МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ИНФОРМАЦИОННОЙ БИЗНЕС-АНАЛИТИКИ

А.В. Мантуленко

Научный руководитель С.П. Борисовна

Бизнес-анализ эволюционировал от статичных отчётов, рассказывающих, что произошло, до интерактивных информационных панелей, с помощью которых вы можете углубиться в данные и попытаться понять, почему это произошло. Новые источники больших данных, включая устройства интернета вещей, подталкивают бизнес перейти от пассивной аналитики — когда мы смотрим на какой-то период в прошлом и ищем тенденции, или раз в день проверяем на наличие проблем — к активной аналитике, способной предупредить о чём-то заранее и позволяющей создавать информационные панели с обновлением в реальном времени. Это помогает лучше использовать операционные данные, которые куда полезнее, если они получены «только что», пока ещё не изменились условия[4].

Машинное обучение заключается в извлечении знаний из данных. Это научная область, находящаяся на пересечении статистики, искусственного интеллекта и компьютерных наук и также известная как прогнозная аналитика или статистическое обучение. В последние годы применение методов машинного обучения в повседневной жизни стало обыденным явлением. Многие современные веб-сайты и устройства используют алгоритмы машинного обучения, начиная с автоматических рекомендаций по просмотру фильмов, заказа еды или покупки продуктов, и заканчивая персонализированными онлайн радиотрансляциями и распознаванием друзей на фотографиях. Когда вы видите сложный сайт типа facebook, amazon или netflix, то весьма вероятно, что каждый раздел сайта содержит несколько моделей машинного обучения [1].

Наиболее успешные алгоритмы машинного обучения – это те, которые

автоматизируют процессы принятия решений путем обобщения известных примеров. В этих методах, известных как обучение с учителем или контролируемое обучение (supervised learning), пользователь предоставляет алгоритму пары объект-ответ, а алгоритм находит способ получения ответа по объекту. В частности, алгоритм способен выдать ответ для объекта, которого он никогда не видел раньше, без какой-либо помощи человека. Если рассмотреть пример классификации спама с использованием машинного обучения, пользователь предъявляет алгоритму большое количество писем (объекты) вместе с информацией о том, является ли письмо спамом или нет (ответы) [1]. Для нового электронного письма алгоритм вычислит вероятность, с которой это письмо можно отнести к спаму.

Алгоритмы обучения без учителя или неконтролируемого обучения (unsupervised algorithms) – это еще один вид алгоритмов, который применяется и имеет свои плюсы [2]. В алгоритмах обучения без учителя известны только объекты, а ответов нет. Есть много успешных сфер применения этих методов, их, как правило, труднее интерпретировать и оценить.

Примерами использования в повседневной жизни данных алгоритмов могут служить следующие задачи:

- определение тем в наборе постов
- сегментирование клиентов на группы с похожими предпочтениями
- обнаружение паттернов аномального поведения на веб-сайте.

Согласно исследованию, проведенному среди 50 000 американских компаний-производителей с 2005 по 2010 годы, количество предприятий, на которых бизнес-решения принимались на основе данных, утроилось. Правда, это всего лишь 30% компаний. А когда в 2015 году телекоммуникационный провайдер Colt провёл опрос среди европейских руководителей IT-компаний, то 71% сказали, что интуиция и личный опыт при принятии решений работают лучше, чем анализ данных (хотя 76% сообщили, что их интуиция не всегда совпадала с полученными ими данными).

Компании-первопроходцы, уже использующие машинное обучение для управления обслуживанием клиентов, финансовыми ресурсами, рисками и соответствием требованиям, как в продажах и маркетинге, так и в новых развиваемых сферах бизнеса, обнаружили «значительное, даже экспоненциальное увеличение выгоды» с точки зрения расходов, выручки и потребительских свойств [3]. Эти компании используют так называемый «перцептивный интеллект» (perceptualintelligence) — комбинацию распознавания речи и голосов, глубокой аналитики и поддержки принятия бизнес-решений.

Способствовать внедрению таких систем будет и вовлечение в их постройку самих бизнес-пользователей. Спрос на специалистов в сфере анализа данных превышает предложение, а это означает, что компании, не имеющие собственных серьёзных разработок, будут обращаться к сторонним аналитическим сервисам. А опытные пользователи (которых в Gartner называют «гражданские специалисты по анализу данных» (citizendatascientists)) будут перенимать эти инструменты и создавать собственные системы глубокой аналитики [4].

Список использованных источников:

1. Андреас Мюллер, Сара Гвидо Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. Вильямс 2017. ISBN978-5-9908910-8-1, 9781449369415
2. Дж. ВандерПлас Python для сложных задач. Наука о данных и машинное обучение. Издательский дом "Питер" 2018. ISBN978-5-4461-0914-2, 978-5-496-03068-7.
3. <https://hbr.org/2016/02/the-rise-of-data-driven-decision-making-is-real-but-uneven>. The Rise of Data-Driven Decision Making Is Real but Uneven.
4. <https://habr.com/ru/company/parallels/blog/318526/>. Машинное обучение как новый инструмент бизнес-анализа.