

УДК 004.8

МАШИННОЕ ОБУЧЕНИЕ – ОСНОВА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В.С. ЛЕСНИЦКИЙ

(Представлено: канд. техн. наук, доц. И.Б. БУРАЧЕНОК)

В статье рассмотрены принципы работы машинного обучения на простых примерах. Актуальность данной темы связана с повсеместным внедрением машинного обучения в разные сферы человеческой жизни, начиная от автоматических роботов-пылесосов, до мобильных телефонов, которые сейчас в 99% случаев не обходятся без машинного обучения. Цель данной статьи объяснить человеку, не связанному с миром ИТ что такое машинное обучение и как оно работает.

Артур Сэмюэл, первопроходец в области искусственного интеллекта и компьютерных игр, ввёл термин «машинное обучение». Он определил машинное обучение как «область обучения, которая даёт компьютерам возможность учиться без явного программирования». Проще говоря, машинное обучение (МО) можно объяснить, как автоматизацию и улучшение процесса обучения компьютеров на основе их опыта без фактического программирования, то есть без какой-либо помощи человека. Процесс начинается с подачи данных хорошего качества, а затем обучения наших машин (компьютеров) путём построения моделей МО с использованием данных и различных алгоритмов. Выбор алгоритмов зависит от того, какой тип данных у нас есть и какую задачу мы пытаемся автоматизировать. [3]

Пример: обучение студентов во время экзамена. При подготовке к экзаменам студенты не заучивают, а стараются выучить его с полным пониманием. Перед экзаменом они загружают свою машину (мозг) большим количеством высококачественных данных (вопросы и ответы из разных книг, заметок учителей или онлайн-видео лекций). Фактически, они тренируют свой мозг как на входе, так и на выходе, то есть какой подход или логика у них есть для решения различных вопросов. Каждый раз, когда они решают практические тестовые задания и верность выполнения (точность/балл), сравнивая ответы, постепенно производительность продолжает расти, приобретая больше уверенности в правильности действий. Вот как на самом деле строятся модели, обучить машину с данными (как входные, так и выходные данные передаются модели), и когда придет время протестировать данные (только с входными данными) и получить оценки нашей модели, сравнив свой ответ с фактическим выходом, который не был загружен во время обучения. Исследователи прилагают все усилия, чтобы улучшить алгоритмы и методы, чтобы эти модели работали ещё лучше. На рисунке 1 показаны отличия работы обычной программы и машинного обучения.



Рисунок 1. – Особенности работы обычной программы и машинного обучения

Далее определим основные отличия МО от традиционного программирования. Традиционное программирование: вводим ДАННЫЕ (ввод) + ПРОГРАММА (логика), запускаем его на машине и получаем вывод. Машинное обучение: вводим данные (ввод) + вывод, вводим их на машине во время обучения, и машина создает свою собственную программу (логику), которую можно оценить во время тестирования.

Что именно означает обучение для компьютера? Считается, что компьютер учится на опыте в отношении некоторого класса задач, если его производительность в данной задаче улучшается с увеличением опыта, а компьютерная программа учится на опыте E в отношении некоторого класса задач T и показателя производительности P , если её производительность в задачах в T , измеренная с помощью P , улучшается с опытом E . [3]

Пример: игра в шашки.

E = опыт игры во многие шашки;

T = задача игры в шашки;

P = вероятность того, что программа выигрывает в следующей игре.

В общем, любую проблему МО можно отнести к одной из двух широких классификаций: обучение с учителем и обучение без учителя. [3]

Посмотрим, как все работает на самом деле. Говоря об онлайн-покупках, есть миллионы пользователей с неограниченным кругом интересов в отношении брендов, цветов, ценового диапазона и многого другого. При совершении покупок в Интернете покупатели обычно ищут несколько товаров. Теперь частый поиск продукта заставит покупателя веб-страницы, поисковую систему или этот интернет-магазин начать рекомендовать или показывать предложения по этому конкретному продукту. Никто не сидит там, чтобы кодировать такую задачу для каждого пользователя, вся эта задача полностью автоматическая. Здесь МО играет свою роль. Исследователи, специалисты по обработке данных, специалисты по МО создают модели на машине, используя хорошее качество и огромное количество данных, и теперь их машина автоматически работает и даже улучшается с увеличением опыта и времени. Традиционно реклама делалась только с использованием газет, журналов и радио, но теперь технологии сделали нас достаточно умными, чтобы делать таргетированную рекламу (система онлайн-рекламы), которая является гораздо более эффективным методом нацеливания на наиболее восприимчивую аудиторию.

В здравоохранении МО также отлично справляется со своей задачей. Исследователи и учёные подготовили модели для обучения машин обнаружения рака, просто глядя на изображения слайд-клеток. Для выполнения этой задачи людям потребовалось бы много времени. Но теперь, больше никаких задержек, машины предсказывают вероятность наличия или отсутствия рака с некоторой точностью, и докторам просто нужно сделать уверенный звонок, вот и все. Ответ на вопрос – как это возможно? – очень прост. Все, что требуется, это высокая вычислительная машина, большой объем данных изображения хорошего качества, модель МО с хорошими алгоритмами для достижения самых современных результатов. Врачи используют МО для диагностики пациентов по различным параметрам.

Ещё один из примеров применения МО. Использование Google Фото, где он распознает лица, Google Lens, где модель распознавания текста МО может извлекать текст из изображений, которые вы загружаете, Gmail, который категоризирует электронную почту как социальную, продвижение, обновления или форум классификация текста, которая является частью МО.

Как работает МО? Сбор прошлых данных в любой форме, подходящей для обработки. Чем выше качество данных, тем больше они подходят для моделирования. Обработка данных. Иногда собранные данные находятся в необработанном виде, и их необходимо предварительно обработать. Пример: в некоторых кортежах могут отсутствовать значения для определенных атрибутов, и в этом случае он должен быть заполнен подходящими значениями для выполнения МО или любой формы интеллектуального анализа данных. Отсутствующие значения для числовых атрибутов, таких как цена дома, могут быть заменены средним значением атрибута, тогда как отсутствующие значения для категориальных атрибутов могут быть заменены атрибутом с наивысшим режимом. Это неизменно зависит от типов используемых нами фильтров. Если данные представлены в форме текста или изображений, тогда потребуется преобразование их в числовую форму, будь то список, массив или матрица. Просто данные должны быть актуальными и последовательными. Он должен быть преобразован в формат, понятный для машины. Разделите входные данные на наборы для обучения, перекрёстной проверки и тестирования. Соотношение между соответствующими наборами должно быть 6: 2: 2. Построение моделей с подходящими алгоритмами и методами на обучающей выборке. Тестирование нашей концептуальной модели с данными, которые не были переданы в модель во время обучения, и оценка её производительности с использованием таких показателей, как оценка F1, точность и отзыв.

Таким образом, машинное обучение работает по принципу человеческого мозга, в самом начале это как новорождённый ребёнок, но чем больше мы его обучаем, тем умнее он становится. Благодаря этому мы можем создавать программы под разные сферы жизнедеятельности человека, и она будет сама обучаться, адаптироваться под конкретного пользователя. Сейчас это очень плотно используется в различных стриминговых сервисах музыки, видео и многих других, для подборки контента для конкретного пользователя.

ЛИТЕРАТУРА

1. Habr.com [Электронный ресурс]. – Режим доступа: <https://www.habr.com>. – Дата доступа: 27.09.2020.
2. Wikipedia.org [Электронный ресурс]. – Режим доступа: <https://ru.wikipedia.org/> – Дата доступа: 27.09.2020.
3. Рашка, Н.С. «Python и машинное обучение: машинное и глубокое обучение с использованием Python, scikit-learn и Ten». / Н.С. Рашка. – М.: Вильямс, 2019 г. – 656 с.