

УДК 004.85

МАШИННОЕ ОБУЧЕНИЕ

Тепляных Д.И., студент гр. ИСт-182, III курс

Научный руководитель: Ерошевич К.В., ассистент кафедры ИиАПС
Кузбасский государственный технический университет имени Т.Ф. Горбачева
г. Кемерово

Введение

Практически каждый современный человек не представляет своей жизни без высоких технологий. Машинное обучение – одна из новейших цифровых отраслей, которая прочно заняла важное место в жизни человечества. Эта технология облегчает труд человека, либо же полностью его заменяет. Машинное обучение даёт возможность людям не работать на опасных для здоровья производствах, помогает анализировать и решать сложные задачи. Например, в горнодобывающей промышленности риск для здоровья и жизни людей очень высок, поэтому внедрение машинного обучения поможет обезопасить и облегчить данную работу, при этом производительность, бесспорно, возрастёт.

Определение и суть машинного обучения

Машинное обучение – современная отрасль технологий, при которой аналитическая система обучается, решая различные задачи, анализируя их. Смысл его в том, что натренированная система сможет принимать решения практически без участия человека, что, безусловно, облегчает жизнь человечества во многих аспектах. Обучение состоит из большого количества алгоритмов и методов, которые при вводе входных данных смогут спрогнозировать какой-либо результат. Вариаций машинного обучения множество: одни системы решают одни задачи, другие решают иные. Очень важно не путаться в терминах машинного обучения, нейросетей и искусственного интеллекта, так как это принципиально разные, хоть и схожие в каких-то моментах, вещи. Искусственный интеллект можно обучить несколькими способами, в том числе методом машинного обучения, из чего следует вывод, что машинное обучение является частью искусственного интеллекта. Из этого можно понять, что машинное обучение – это обучение программы, в ходе которого она постепенно улучшает метод выполнения поставленной цели, и включает в себя огромное количество различных вычислительных, математических, статистических методов разработки алгоритмов, которые помогают решить задачу, ища закономерности в разных входных данных. Программа получает решение не по формуле, а по

зависимости результатов от других различных наборов входных данных и значений. Отсюда можно сделать вывод, что главный смысл машинного обучения – научить машину «учиться», и искать метод решения самой, вычлняя из прошлых решений полезные для неё знания.

Задачи машинного обучения

Идея машинного обучения основывается на том, что системы и программы смогут научиться принимать решения, выявлять закономерности без помощи человека.

В машинном обучении принято выделять четыре основных задачи:

- **регрессия** — предположение числовых значений чего-либо, основываясь на прошлых данных, например, прогнозирование будущих объемов продаж какой-либо компании, опираясь на результаты прошлых продаж;
- **классификация** — определение класса объекта, основываясь на прошлых знаниях, например, определение болезни человека по фотографии, опираясь при этом на фотографии других больных людей;
- **кластеризация** — распределение входных данных на категории, например, разделение базы потребителей по признаку платёжеспособности;
- **снижение размерности** — объединение, сжатие большого количества признаков к нескольким, для того чтобы выявить главную информацию, получая при этом практически тот же результат, например, сжатие файла;
- **выявление аномалий** — поиск необычных, незнакомых для программы вещей, которые не совпадают с основной массой данных, например, выявление мошеннических переводов.

Способы и методы машинного обучения

Способы машинного обучения обычно разделяют на три вида:

Обучение с учителем - для программы изначально задаётся большое количество входных данных с уже готовыми с результатами, на основании которых она сможет сделать какой-либо вывод.

Обучение без учителя – программе даются только входные данные без результатов, и ей предстоит найти сходства и зависимости в них.

Глубокое обучение – При этом виде машинного обучения используются два метода вместе: с учителем, и без. Программе задаются все входные данные, но результаты известны не на все. Это позволяет разделить большую задачу на несколько маленьких, при помощи делегирования их другим устройствам. К

примеру, процессор, который получил информацию, передает её другим процессорам, те ее анализируют и передают далее.

Методы машинного обучения

Чаще всего, в наше время, для разработки программ и систем машинного обучения люди используют языки Python, R, Scala, Julia, так как они поддерживаются практически всеми средами разработки, например, VisualStudio, R-Studio, Eclipse, Spyder, Juno и др.

Нейронные сети - данный метод очень напоминает структуру мозга, а именно, нейронов – цифровой нейрон, также как и человеческий, имеет связь с несколькими другими нейронами, тем самым информация может передаваться в любую точку сети. В итоге, данные выходят на верхний уровень, где сеть предлагает решение поставленной ей задачи.

Дерево решений – данный алгоритм распределяет объекты по группам, основываясь на их признаках, и в зависимости от ответа выбирается все меньшая и меньшая группа, (сук, потом ветка) пока ответ не будет «листом» то есть, конечным.

«Случайный лес» - дерево решений работает только тогда, когда оно уже хоть минимально обучено, в отличии от метода «случайный лес». Данный алгоритм использует для принятия решения случайно созданные «деревья», и даёт им шанс «проголосовать». Таким образом, этот алгоритм является одним из самых быстрых в обучении, а также универсальным.

Кластеризация – данный алгоритм группирует объекты, которые имеют какие-либо похожие качества, при помощи статистических алгоритмов. Он является самым удобным при решении задач классификации.

Поиск ассоциативных правил – данный алгоритм находит отношения и связи между своими данными, тем самым связывая их «ассоциациями». Одним из популярных примеров работы этого метода является то, с чем сталкивались практически все пользователи интернета: при заказе какого-либо товара на сайте, появляется надпись: «Вместе с этим продуктом часто приобретают...».

Алгоритмы машинного обучения

Алгоритмы машинного обучения всегда включают в себя обучение функции f , которая анализируют входные данные – x , и выводит переменную y . Прогностическое моделирование – это прогнозирование значений y для известных значений x , и это является самой популярной задачей в машинном обучении, при этом его цель – сделать результат наиболее точным.

Алгоритмов машинного обучения довольно много, расскажу про некоторые из них:

Линейная регрессия — алгоритм машинного обучения, который наиболее простой в использовании по сравнению с другими, и, следовательно, самый известный. В этом алгоритме взаимосвязь между переменными обозначается линейно – то есть прямой.

Логистическая регрессия — этот алгоритм появился из раздела статистики, и именно он очень хорош для задач, где нужно разделить на две группы какие-либо объекты или данные.

LDA – когда нужно разделить объекты более, чем на две группы, то больше подойдет алгоритм LDA. Этот алгоритм состоит из статистических признаков данных, которые рассчитываются отдельно для каждой группы.

Дерево решений – данный алгоритм можно изобразить как двоичное дерево, которое знакомо многим по урокам информатики в школе. Узлы дерева являются точкой деления для входной переменной.

«Наивный» Байесовский классификатор - довольно простой для понимания алгоритм, который удивляет своей эффективностью. Забавное слово «наивный» в названии говорит о том, что алгоритм предполагает, что любая переменная является независимой, что не соответствует реальным данным. Алгоритм состоит из вероятностей, которые рассчитываются с помощью тренировок: вероятность каждой группы, и условная вероятность для каждой группы при любом значении переменной x .

Гиперплоскость — этот алгоритм разделяет пространство входных данных и переменных линией, а разделяет он их на ноль и один. Пока алгоритм обучается, он ищет коэффициенты, которые на начальных этапах помогают ему еще точнее разделять входные данные линией.

Бэггинг и рандомный лес – еще один популярный алгоритм, который является отраслью ансамблевого алгоритма, называемый Беггинг.

Бустинг и AdaBoost – это ансамблевые алгоритмы, при которых программа создается, учитывая ошибки прошлой – создается алгоритм, затем еще один, который пытается исправить ошибки прошлого. Этот цикл продолжается до того момента, пока результат программы не будет максимально точным.

K-ближайших соседей (KNN) - легкий и довольно полезный алгоритм, который хранит в себе огромный набор тренировочных данных.

Сети векторного квантования (LVQ) – данный алгоритм практически такой же как KNN, за исключением одного – ему не нужно хранить в памяти весь тренировочный набор данных.

Метод опорных векторов (SVM) – довольно популярный алгоритм машинного обучения.

Примеры реального использования машинного обучения

Приведу несколько примеров использования машинного обучения в реальной жизни:

Google — на данный момент практически нереально найти какую-либо сферу технологий, в которую не внесла вклад данная компания. В последние года Google разрабатывали технологии, которые могут позволить замедлить старение, разрабатывали нейронные сети и различные медицинские устройства. Эта компания активно изучает и развивает тему машинного обучения, что, бесспорно, выдвигает компанию с каждым годом на новый уровень. Улучшение ранжирования, предсказательные системы, развитие классических алгоритмов, увеличение эффективности перевода и распознавания текста – всё это заслуга Google.

Facebook — социальная сеть, которая активно использует искусственный интеллект для очистки ленты от спама и непристойного материала. Также, у данной сети есть FacebookMessenger – это «чатбот», который разработан настолько хорошо, что не всегда можно понять, сидит ли по ту сторону экрана человек, или нейросеть. Благодаря этому сервису компании, которые будут ее использовать, смогут предлагать своим потенциальным покупателям проконсультироваться с «чатботом», который ответит на вопросы клиента.

Twitter — социальная сеть, которая представляет собой новостную ленту, тоже активно пользуется машинным обучением. Теперь лента выстраивается алгоритмом на основе интересов человека, что, несомненно, понравится всем пользователям, и повысит рейтинг данной социальной сети. Искусственный интеллект Twitter оценивает каждый «твит» по своим критериям, благодаря которым может показать его в ленте тем, кто его гипотетически оценит.

Baidu — китайская поисковая система тоже проявляет активное участие в развитии искусственного интеллекта. Одной из самых необычных и интересных разработок компании является нейронная сеть под названием DeepVoice. Она может синтезировать голоса людей, которые очень сложно отличить от настоящих, так как она филигранно имитирует акценты, высоту голоса, различные интонации и тд. Последняя версия данной разработки очень поможет для более эффективного распознавания речи, что объективно облегчит быт пользователям интернета.

IBM — известная большинству компания поменяла свою сферу деятельности, и на данный момент довольно активно осваивает новые направления. Watson – самый известный продукт бренда. Это искусственный интеллект, который эффективно диагностирует различные виды рака, тем самым облегчая работу онкологов.

Заключение

Исходя из всего вышенаписанного, можно сделать вывод, что машинное обучение – это отрасль искусственного интеллекта, которая не сможет работать без знаний математики, статистики, теории вероятностей. Она способна обучаться, опираясь на свой прошлый, неудачный опыт.

Машинное обучение уже оказывает на нашу жизнь огромное влияние, несмотря на то, что эта отрасль программирования появилась сравнительно недавно. С его помощью, например, людям подбирают по их вкусам и интересам музыку и фильмы, даже рекламу. Без машинного обучения было бы намного сложнее искать нужную нам информацию.

В ближайшем будущем уже каждый человек сможет применять машинное обучение для своих нужд, так как эта сфера по сей день активно развивается, и становится проще и совершеннее, дабы каждому обывателю интернета.

Список литературы:

1. 17 примеров применения машинного обучения в 5 отраслях бизнеса [Электронный ресурс] // Яндекс Дзен – Режим доступа: <https://zen.yandex.ru/media/mcs/17-primerov-primeneniia-mashinnogo-obucheniia-v-5-otrasliah-biznesa-5f60a7d8249b32282b9c47ef/> – (Дата обращения: 01.02.2021).
2. Что такое машинное обучение? Методы, типы, задачи и примеры машинного обучения [Электронный ресурс] // Майнинг криптовалюты: информационный аналитический портал – Режим доступа: <https://mining-cryptocurrency.ru/mashinnoe-obuchenie-metody-tipu/> – (Дата обращения: 01.02.2021).
3. Машинное обучение [Электронный ресурс] // Я неуч! – Режим доступа: https://www.yaneuch.ru/cat_28/mashinnoe-obuchenie/329035.2460630.page1.html – (Дата обращения: 01.02.2020).
4. Машинное обучение – это легко [Электронный ресурс] // Хабр:Сообщество IT- специалистов – Режим доступа: <https://habr.com/ru/post/319288/> – (Дата обращения: 11.03.2021).
5. Применение машинного обучения и Data Science в промышленности [Электронный ресурс] // Хабр: Сообщество IT- специалистов – Режим доступа: <https://habr.com/ru/company/mailru/blog/462769/> – (Дата обращения: 11.03.2021).

6. Обзор задач компьютерного зрения в медицине [Электронный ресурс]
// Хабр: Сообщество IT- специалистов – Режим доступа:
<https://habr.com/ru/post/309152/> – (Дата обращения: 12.03.2021).