

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО РЫБОЛОВСТВУ
АССОЦИАЦИЯ ГОСУДАРСТВЕННЫХ УНИВЕРСИТЕТОВ ПРИКАСПИЙСКИХ СТРАН
ФГБОУ ВО «АСТРАХАНСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»



70-я МЕЖДУНАРОДНАЯ
СТУДЕНЧЕСКАЯ
НАУЧНО-ТЕХНИЧЕСКАЯ
КОНФЕРЕНЦИЯ,
ПОСВЯЩЕННАЯ 90-ЛЕТИЮ
АИРХ-АТИРПиХ-АГТУ

Астрахань, 13–18 апреля 2020 года

Материалы

Астрахань

Издательство АГТУ 2020

© ФГБОУ ВО «Астраханский государственный технический университет»

Е. П. Алемасов, Р. С. Зарипова

Казанский государственный энергетический [На главную](#) [Предыдущий](#) университет,
[Следующий](#)

Казань, Российская Федерация

zarim@rambler.ru

АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ И ИХ ИСПОЛЬЗОВАНИЕ

В данной статье рассматриваются наиболее популярные алгоритмы машинного обучения и их использование в повседневной жизни.

Машинное обучение (МО) определяется как процесс, в результате которого компьютеры способны показать поведение, которое в них не было явно запрограммировано. В настоящее время МО, наука о данных и искусственный интеллект становятся всё более популярными, поскольку они могут быть полезны во всех сферах жизнедеятельности [1]. МО может извлекать значимые данные путем обнаружения скрытых образцов из данных. Благодаря алгоритмам МО у компьютера есть возможность учиться и повышать свою эффективность. Существует множество данных, которые с помощью МО легко превращаются в знания [2].

Алгоритмы МО обычно можно разделить на две основные группы: контролируемые и неконтролируемые. Для контролируемых алгоритмов требуется специалист по данным, обладающий знаниями в области МО, для обеспечения правильного ввода. Учитель решает, какие функции модель должна учитывать и использовать для разработки прогнозов. Неконтролируемые алгоритмы обучения обычно используются для решения сложных задач обработки по сравнению с контролируемыми системами обучения, включая распознавание изображений, преобразование речи в текст и создание естественного языка. Эти нейронные сети работают, комбинируя миллионы примеров, обучающих данных и автоматически находя корреляции между элементами.

Основная цель контролируемых алгоритмов состоит в том, чтобы извлечь уроки из обучающего набора данных, затем сделать прогноз и получить желаемый результат. Основную долю задач контролируемого обучения составляют классификация и регрессия. В случае регрессии алгоритм возвращает числовое целевое значение для каждого экземпляра,

например, какой доход будет от новой маркетинговой кампании. Линейная регрессия является самым популярным и известным контролируемым алгоритмом. Прогнозирующее моделирование связано с минимизацией ошибки модели или созданием максимально точных прогнозов. В данном случае заимствуются и повторно используются алгоритмы из разных областей, включая статистику.

Логистическая регрессия – это функция, значением которой является вероятность принадлежности к определенному классу, использующая одну модель полиномиальной логистической регрессии с одной оценкой. Логистическая регрессия аналогична линейной регрессии в том смысле, что цель состоит в том, чтобы найти значения для индексов, которые взвешивают каждую входную переменную.

Деревья решений – это иерархические структуры, позволяющие классифицировать входные данные или предсказывать выходные значения по заданным исходным значениям. Листовые узлы дерева содержат выходную переменную, которая является выходной категорией. Прогнозы делаются путем обхода расщеплений дерева до узлов.

Наивный байесовский алгоритм – это алгоритм, используемый для решения задач классификации. Он использует теорему Байеса с предположением независимости между предикторами. Байесовский классификатор предполагает, что существование определенного признака в классе не связано с наличием другого признака. Модель этого алгоритма проста в построении и полезна при работе с большими наборами данных.

Алгоритм К-ближайших соседей используется для решения задач классификации. Это простой алгоритм, который хранит все доступные случаи и классифицирует новые случаи большинством голосов своих k соседей. Он может потребовать много места для хранения всех данных.

Неконтролируемое обучение – это класс методов МО для поиска шаблонов в наборах данных. Данные, переданные неконтролируемому алгоритму, не помечены. То есть даны только входные переменные без выходных переменных. При таком обучении алгоритмы предоставлены самим себе для обнаружения интересующих структур в данных.

Кластеризация К-средних – нахождение группы в данных с количеством категорий, представленных переменной K . Этот алгоритм работает итеративно, чтобы назначить каждую точку данных одной из K групп на основе предоставленных функций. Точки данных сгруппированы на основе сходства признаков. Этот алгоритм используется для поиска групп, которые не были явно помечены в данных.

Карта самоорганизации (SOM) – это искусственная нейронная сеть, основанная на обучении без учителя. SOM может использоваться для обнаружения особенностей, присущих проблеме, поэтому ее также называют картой самоорганизующейся функции. В картах самоорганизации нейроны помещены в узлах решетки.

Приведем некоторые примеры использования алгоритмов МО.

Онлайн рекомендации. Машинное обучение позволяет ритейлерам рекомендовать пользователям индивидуальные предложения на основе их предыдущих покупок или активности (поиск Google, Amazon и т. д.).

Лучшее обслуживание клиентов и системы доставки. В компаниях, где время отклика ограничено кадровыми ресурсами, интеллектуальные машины могут расшифровать смысл электронных писем и накладных, чтобы расставить приоритеты для задач, для облегчения нагрузки [3].

Отслеживание изменений цен. Цена розничных товаров имеет тенденцию колебаться в течение определенного периода времени. МО помогает компаниям электронной торговли отслеживать закономерности этих колебаний и устанавливать цены в соответствии со спросом [4].

Системы распознавания голоса (Siri, Алиса) используют МО и глубокие нейронные сети для имитации взаимодействия человека. Далее эти приложения научатся «понимать» нюансы и семантику нашего языка.

Google Maps анализирует скорость трафика с помощью анонимных данных о местоположении со смартфона. Используя такие данные, Google может предложить самые быстрые маршруты. PayPal использует алгоритмы МО для обнаружения и борьбы с мошенничеством. Внедряя методы глубокого обучения, PayPal может анализировать большое количество данных о клиентах и оценивать риски. Использование МО всё больше охватывает нашу повседневную жизнь. Мы ежедневно пользуемся благами алгоритмов МО, которые стали частью нашего существования.

Список литературы

1. Пырнова О.А., Зарипова Р.С. Перспективы развития искусственного интеллекта и кибернетики // Информационные технологии в строительных, социальных и экономических системах. 2019. № 3-4 (17-18). С. 78-81.
2. Пырнова О.А., Зарипова Р.С. Технологии искусственного интеллекта в образовании // Преподавание информационных технологий в РФ: Материалы Семнадцатой открытой Всеросс. конф. 2019. С. 335-337.
3. Басаргин В.Я., Зарипова Р.С., Пырнова О.А. Влияние цифровых технологий на урбанизацию // «Цифровая культура открытых городов»: Сборник материалов Международной научно-практической конференции. Екатеринбург. 2018. С. 13-15.

4. Зарипова Р.С., Галямов Р.Р., Шарифуллина А.Ю. Организация производства в условиях цифровой экономики // Наука Красноярья. 2019. Т.8. № 1-2. С. 20-23.