

ТЕХНИЧЕСКИЕ НАУКИ

УДК 004.021

Акжолов Р.К.

студент 3 курса факультета математики и информационных технологий АГУ, кафедра информационных технологий.

Астраханский государственный университет
(Россия, г. Астрахань)

МАШИННОЕ ОБУЧЕНИЕ

Аннотация: в статье проведён анализ методов машинного обучения. Рассмотрены несколько методов, приведены примеры алгоритмов обучения с учителем и без учителя, а также описаны сферы применения.

Ключевые слова: машинное обучение, нейронные сети, кластеризация, регрессия, классификация, алгоритм, данные.

Машинное обучение – это категория алгоритмов, которая позволяет программным приложениям быть более точными в прогнозировании результатов без явного программирования. Основная предпосылка машинного обучения заключается в создании алгоритмов, которые могут принимать входные данные и использовать статистический анализ для прогнозирования выходных данных по мере появления новых данных.

Процессы, участвующие в машинном обучении, похожи на процессы интеллектуального анализа данных и прогнозного моделирования. И то, и другое требует поиска данных для поиска шаблонов и соответствующей корректировки действий программы. Многие люди знакомы с машинным обучением по интернет-магазинам и получаемой от них рекламе. Это происходит потому, что рекомендательные системы используют машинное обучение для персонализации доставки онлайн-рекламы практически в реальном времени. Помимо персонализированного маркетинга, другие распространенные случаи использования

машинного обучения включают обнаружение мошенничества, фильтрацию спама, обнаружение угроз сетевой безопасности, профилактическое обслуживание и создание новостных лент.

Как работает машинное обучение? Алгоритмы машинного обучения классифицируются на обучение с учителем и без учителя. Для обучения с учителем требуется, чтобы специалист по данным обладал навыками машинного обучения, чтобы обеспечить как ввод, так и желаемый результат, в дополнение к предоставлению обратной связи о точности предсказаний во время обучения алгоритму. Исследователи данных определяют, какие переменные или характеристики модель должна анализировать и использовать для разработки прогнозов. После завершения обучения алгоритм будет применять полученные данные к новым данным. Примерами методов обучения с учителем являются классификация и регрессия.

Классификация – самая популярная задача машинного обучения. Цель этого метода классифицировать объекты по заранее известному признаку, например, документы по языкам, музыку по жанрам и т.д. Классическим примером классификации является алгоритм «Наивный Байес». Раньше он применялся в спам-фильтрах, сейчас этим занимаются деревья решений. Другой пример классификации – метод опорных векторов, он является самым популярным методом классификации. Раньше им классифицировали все: виды растений, человеческие лица на фотографиях, документы по темам.

Регрессия используется в задачах, когда надо предсказать результат. Регрессия задается в виде уравнения, которое описывает функцию, показывающую взаимосвязь между входными и выходными данными. Этот метод машинного обучения используется в: прогнозировании стоимости ценных бумаг, медицинских диагнозах и т.д. Популярные алгоритмы регрессии – линейная, полиномиальная, логистическая и т.д.

Методы обучения без учителя не требуют готовых ответов и используются для поиска зависимостей между объектами. К таким методам относятся: кластеризация, снижение размерности и нейронные сети.

Кластеризация – метод машинного обучения, при котором объекты разделяются по неизвестному признаку, при этом алгоритм сам выделяет эти признаки. На сегодняшний день этот метод нашёл своё применение в таких задачах, как сжатие изображений, объединение близких точек на карте и т.д. Наиболее популярными алгоритмами являются метод К – средних и DBSCAN.

Метод снижения размерности используется для того, чтобы собрать признаки в абстракции более высокого уровня. Например, в каких-нибудь исследованиях часто требуется визуализировать данные для удобного изучения и анализа. Так, алгебраические функции могут быть представлены в виде графиков функций, статистические выборки – в виде диаграмм рассеивания. Трудности появляются, когда необходимо визуализировать многомерные данные, когда каждый объект имеет много признаков. Графически такие объекты представляются как точка в многомерном пространстве. Человеку удобно анализировать данные, отображенные на плоскости, поэтому применяют алгоритмы снижения размерности. Примерами таких алгоритмов являются: метод главных компонент, сингулярное разложение, алгоритм T-SNE. На сегодняшний день эти алгоритмы используются в рекомендательных системах, тематическом моделировании и т.д.

Нейронные сети используются для более сложных задач обработки, чем контролируемые системы обучения, включая распознавание изображений, преобразование речи в текст и создание естественного языка. Эти нейронные сети работают, комбинируя миллионы примеров обучающих данных и автоматически выявляя тонкие корреляции между многими переменными. После обучения алгоритм может использовать свой банк ассоциаций для интерпретации новых данных. Эти алгоритмы стали возможными только в эпоху больших данных, поскольку они требуют огромных объемов обучающих данных.

Список литературы:

1. Машинное обучение для людей // Ватрик ру URL: https://vas3k.ru/blog/machine_learning/

2. Обзор самых популярных алгоритмов машинного обучения // Tproger URL:
<https://tproger.ru/translations/top-machine-learning-algorithms/>

3. Машинное обучение // MachineLearning.ru URL:
http://www.machinelearning.ru/wiki/index.php?title=Машинное_обучение

4. Машинное обучение для чайников // Newtonew URL:
<https://newtonew.com/tech/machine-learning-novice>