

ВВЕДЕНИЕ

Трудно переоценить вклад, который внесло развитие машинного обучения в современный мир.

На сегодняшний день технологии машинного обучения успешно применяются практически во всех сферах деятельности человека: наука, коммерция, медицина. Мы пользуемся результатами работы алгоритмов машинного обучения каждый день даже не подозревая об этом. Системы распознавания рукописного ввода, голоса и изображений доступны для использования любому человеку. Беспилотные самолеты и автомобили уже не вызывают ни у кого удивления. Пользуясь интернетом каждый день мы сами того не подозревая принимаем участие в обучении разного рода алгоритмов – поисковые системы, контекстная реклама, рекомендации в интернет магазинах анализируют данные, полученные от пользователей для повышения эффективности работы сайтов и увеличения продаж.

Первые научные работы в данном направлении появились еще более 60 лет назад. В 1952 г. Артур Самуэль представил программу Checkers-Playing для компьютера IBM 701, в которой имелась возможность самообучения[1]. Это считается первым примером использования машинного обучения. Но до начала первого десятилетия XXI века эта сфера развивалась относительно медленно и не интересовала практически никого кроме ученых. Взрывной рост популярности подходов машинного обучения обусловлен следующими факторами:

1. Большие данные. Повсеместное внедрение компьютерных технологий, а также развитие Интернета привели к тому что различные организации получили возможность накапливать огромные (сотни и тысячи терабайт) объемы цифровых данных. Данные о банковских транзакциях, медицинские карты, данные о погоде, полученные с метеорологических станций - для работы с таким количеством информации традиционные подходы стали неприменимы. Поэтому разработка новых методов анализа данных была вызвана не интересом исследователей и ученых, а практической необходимостью.

2. Быстрое снижение стоимости компьютерной памяти и параллельных вычислений. В 2004 г. компания Google раскрыла свою технологию MapReduce, за которой следовал выход ее открытого аналога Hadoop (2006 г.) Возможность быстро обрабатывать большие объемы данных при помощи параллелизации еще больше подтолкнуло развитие машинного обучения.

Одним из наиболее известных методов машинного обучения являются нейронные сети – модели, построенные по принципу и организации биологических нейронных сетей, т.е. сетей нервных клеток живого организма. Нейронные сети возникли как попытка смоделировать процессы, происходящие в мозге живого организма. Сегодня различные виды нейронных сетей широко используются для решения задач распознавания образов, кластеризации, классификации и др.

Все задачи, которые решаются с помощью машинного обучения можно разделить на 2 основных типа:

1. Обучение с учителем. При таком типе задач происходит обучение системы с помощью набора примеров, (обучающей выборки). Другими словами известен набор входных данных системы и соответствующих им эталонных выходных данных. Необходимо восстановить зависимость между входными и выходными данными, то есть построить алгоритм который способен для любого нового набора входных данных выдать более или менее точный ответ.

2. Обучение без учителя. При этом системе предоставляется массив данных, в которыми необходимо найти определенные закономерности, например сгруппировать множества объектов таким образом, чтобы поместить максимально похожие между собой элементы в одну группу (задача кластеризации).

В результате на данный момент нам доступен широкий спектр алгоритмов, каждый из которых обладает своими достоинствами и недостатками. Поэтому для каждой конкретной проблемы необходимо проанализировать все факторы которые могут повлиять на эффективность полученного решения, так как неправильно принятые решения на этапе выбора конкретного алгоритма, его конфигурации могут привести к значительным временным и финансовым потерям в будущем. В существующих программных средствах для работы с машинным обучением обычно имеется возможность получать доступ к информации о функционировании системы для того чтобы выявлять ошибки еще на этапе проектирования.

Целью дипломного проекта является реализация программного комплекса предоставляющего средства для решения проблем машинного обучения с помощью нейронных сетей, а также позволяющего осуществлять диагностику полученной системы при помощи визуализации и сбора данных о процессе работы.