

Таблица 1

n	0	1	2	3	4	5
m	0	4	8	20	26	34
l	0	4	4	12	6	8
r	0	1	1	1,414	1,732	...

Таблица 2

n	0	1	2	3	4	5
m	0	12	18	42	54	78
l	0	12	6	24	12	24
r	0	0,707	1	1,225	1,414	...

ЛИТЕРАТУРА

1. Ландау Л. Д., Лифшиц Е. М. Статистическая физика. – М: Наука, 1976. – 584 с.
2. Наркевич И. И, Волмянский Э. И., Лобко С. И. Физика. – Минск: Новое знание, 2004. – 679 с.

УДК 004.588

Магистрант О.Л. Панченко
 Науч. рук. доц. Н.В. Пацей
 (кафедра программной инженерии, БГТУ)

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ

Искусственный интеллект – это раздел информатики, посвященный моделированию интеллектуальной деятельности человека. В настоящее время искусственный интеллект открывает перспективы, без которых дальнейшее развитие цивилизации немыслимо. Уже стало обыденным то, что компьютеры «умнеют» буквально на глазах, а компьютерные программы становятся все более и более интеллектуальными. Сейчас понятие обучение уже применимо не только к человеку, но и к машине. Об этом и пойдет речь. А именно о машинном обучении, о том как оно происходит.

Машинное обучение (Machine Learning) – обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться. Существует два типа обучения. Обучение по прецедентам, или индуктивное обучение, основано на выявлении общих закономерностей по частным эмпирическим данным. Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний. Его принято относить к области экспертных систем.

Практически ни одно исследование в машинном обучении не обходится без эксперимента на модельных или реальных данных, подтверждающего практическую работоспособность метода.

Общая постановка задачи обучения по прецедентам – дано конечное множество прецедентов, по каждому из которых собраны некоторые данные. Данные о прецеденте называют также его описанием. Совокупность всех имеющихся описаний прецедентов называется обучающей выборкой. Необходимо по этим частным данным выявить общие зависимости, взаимосвязи, присущие не только этой конкретной выборке, но вообще всем прецедентам, в том числе тем, которые ещё не наблюдались.

Далее вводится функционал качества. Это значение показывает, насколько хорошо модель описывает наблюдаемые данные. Алгоритм обучения (learning algorithm) ищет такой набор параметров модели, при котором функционал качества на заданной обучающей выборке принимает оптимальное значение.

Типология задач обучения по прецедентам:

— обучение с учителем (supervised learning) – каждый прецедент представляет собой пару «объект, ответ»;

— обучение без учителя (unsupervised learning). В этом случае ответы не задаются, и требуется искать зависимости между объектами;

— обучение с подкреплением (reinforcement learning). Роль объектов играют пары «ситуация, принятое решение».

Выделяют несколько типов функционалов качества:

— при обучении с учителем — функционал качества может определяться как средняя ошибка ответов. Предполагается, что искомым алгоритм должен его минимизировать. Для предотвращения переобучения в минимизируемый функционал качества часто в явном или неявном виде добавляют регуляризатор;

— при обучении без учителя — функционалы качества могут определяться по-разному, например, как отношение средних межкластерных и внутрикластерных расстояний;

— при обучении с подкреплением — функционалы качества определяются физической средой, показывающей качество приспособления агента.

Целью машинного обучения является частичная или полная автоматизация решения сложных профессиональных задач в самых разных областях человеческой деятельности.

Машинное обучение имеет широкий спектр приложений: распознавание речи, жестов, рукописного ввода; техническая и медицинская диагностика; прогнозирование временных рядов; обнаружение мошенничества; обнаружение спама; категоризация документов и т. д. Сфера применений машинного обучения постоянно расширяется.

Одним из наиболее распространенных инструментов Data Scientist'a на сегодняшний день является Python. Он выбран по например таким причинам как: предельная простота языка и набор готовых библиотек. У Pythona есть такой дистрибутив как Anaconda, который включает в себя набор библиотек для научных и инженерных расчетов. Наиболее применяемые такие как:

— NumPy – это библиотека, которая позволяет Python быстро обрабатывать данные;

— высокоуровневая библиотека Pandas предназначена для анализа данных;

— дополнение к библиотеки Pandas, пакет matplotlib – возможность рисовать графики к полученным данным;

— Scikit-Learn – библиотека Python с алгоритмами машинного обучения [3].

Далее будет описаны шаги работы со Scikit-Learn.

1. Загрузка данных. В первую очередь данные необходимо загрузить в оперативную память, чтобы мы имели возможность работать с ними. Сама библиотека Scikit-Learn использует в своей реализации NumPy массивы;

2. Нормализация данных. Большинство градиентных методов (на которых по-сути и основаны почти все алгоритмы машинного обучения) сильно чувствительны к шкалированию данных. Поэтому перед запуском алгоритмов чаще всего делается либо нормализация, либо так называемая стандартизация. Нормализация предполагает замену номинальных признаков так, чтобы каждый из них лежал в диапазоне от 0 до 1.

3. Отбор признаков. Зачастую самым важным при решении задачи является умение правильно отобрать и даже создать признаки. В англоязычной литературе это называется Feature Selection и Feature Engineering. В то время как Feature Engineering довольно творческий процесс и полагается больше на интуицию и экспертные знания, для Feature Selection есть уже большое количество готовых алгоритмов;

4. Построение алгоритма. Как уже было отмечено, в Scikit-Learn реализованы все основные алгоритмы машинного обучения. Рассмотрим некоторые из них.

5. Оптимизация параметров алгоритма. Одним из самых сложных этапов в построении действительно эффективных алгоритмов является выбор правильных параметров. Обычно, это делается легче с опытом, но так или иначе приходится делать перебор. К счастью, в Scikit-Learn уже есть немало реализованных для этого функций

Мы рассмотрели весь процесс работы с библиотекой Scikit-Learn. Заключительным этапом является вывод результатов обратно в файл [4].

Python – это мощный и лаконичный язык программирования, который теперь стал одной из основных частей моего инструментария разработчика. Он прекрасно подходит для реализации машинного обучения, обработки естественного языка и математических и научных приложений.

ЛИТЕРАТУРА

1. Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell (1983), *Machine Learning: An Artificial Intelligence Approach*, Tioga Publishing Company, ISBN 0-935382-05-4.

2. Машинное обучение. [Электронные ресурс] – Режим доступа: http://wreferat.baza-referat.ru/Машинное_обучение (дата обращения 05.03.2018)

3. 27 шпаргалок по машинному обучению и Python в 2017. [Электронные ресурс] – Режим доступа: <https://proglib.io/p/ds-cheatsheets/> (дата обращения 06.03.2018)

4. Введение в машинное обучение с помощью Python и Scikit-Learn. [Электронные ресурс] – Режим доступа: <https://habrahabr.ru/company/mlclass/blog/247751/> (дата обращения 05.03.2018)

УДК 004.934.5

Студ. В.В. Ильин
Науч. рук. доц. А.С. Кобайло
(кафедра программной инженерии, БГТУ)

ЭЛЕКТРОННЫЙ АРХИВ ДОКУМЕНТОВ ДЛЯ ERP СИСТЕМЫ MICROSOFTDYNAMICSAX

Мало кто станет отрицать важность и значимость процессов, день ото дня движущих предприятиями и организациями, в которых мы работаем и с которыми взаимодействуем. Бизнес-процессы представляют собой ключевые действия, в результате выполнения которых