

Имплементация машинного обучения на веб-платформе на примере проектирования рекомендательных систем

Студент: Костюченко Ю.А.

Научный руководитель: Государев И.Б., к.п.н., доцент



Актуальность

Потребность в рекомендательных системах и эффективных алгоритмах для них



Объект

Рекомендательная система, разработанная с использованием технологий машинного обучения



Предмет

Имплементация алгоритмов, используемых для создания персональных рекомендаций



Цель

Создание рекомендательной системы фильмов на веб-платформе с использованием алгоритмов машинного обучения



Задачи

- изучить категориальный аппарат машинного обучения и рекомендательных систем;
- провести обзор научных источников отечественных и зарубежных авторов по теме исследования;
- исследовать имеющиеся на рынке аналоги или похожие по функционалу системы, выделить их достоинства и недостатки;
- выбрать тип разрабатываемой системы;
- сравнить существующие виды алгоритмов и выбрать наиболее подходящий и быстрый;
- выполнить моделирование основной части системы;
- имплементировать необходимые для полноценной работы системы алгоритмы.



Теоретическая значимость

Все технологии, методы и алгоритмы, а также их комбинирование, будут подробно описаны и представлены научному сообществу, в результате чего с их помощью могут быть проведены новые исследования.



Практическая значимость

В результате выполнения работы будут предложены наиболее эффективные технологии и алгоритмы, которые необходимо использовать для создания подобной системы. Это позволит ускорить уже существующие системы, а также способствовать возникновению более эффективных рекомендательных систем.



Машинное обучение

Обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться.

Находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения.



Методы машинного обучения без учителя

Неконтролируемые преобразования (unsupervised transformations) – это алгоритмы, создающие новое представление данных, которое в отличие от исходного представления человеку или алгоритму машинного обучения будет обработать легче. Общераспространенное применение неконтролируемых преобразований – сокращение размерности. Мы берем высокоразмерное представление данных, состоящее из множества признаков, и находим новый способ представления этих данных, обобщая основные характеристики и получая меньшее количество признаков. Общераспространенное применение сокращения размерности – получение двумерного пространства в целях визуализации.



Методы машинного обучения без учителя

Алгоритмы кластеризации (clustering algorithms) разбивают данные на отдельные группы схожих между собой элементов. Рассмотрим пример загрузки фотографий в социальной сети. Часто вы формируете запросы типа «покажите мне все фотографии, на которых изображен Иван Петров». Для выполнения подобных запросов, администрация сайта, возможно, захочет сгруппировать фотографии, на которых изображен один и тот же человек. Однако при этом неизвестно, 148 на каких загружаемых фотографиях кто показан, и неизвестно, какое количество различных пользователей присутствует на ваших фотографиях. Разумный подход заключался бы в том, чтобы извлечь все лица и разделить их на группы лиц, которые схожи между собой. Будем надеяться, что они соответствуют одному и тому же человеку и изображения в сгруппированном виде будут предъявлены.



Методы машинного обучения с учителем

Задача классификации (classification) отличается тем, что множество допустимых ответов конечно. Их называют метками классов (class label). Класс — это множество всех объектов с данным значением метки.

Задача регрессии (regression) отличается тем, что допустимым ответом является действительное число или числовой вектор.

Задача ранжирования (learning to rank) отличается тем, что ответы надо получить сразу на множестве объектов, после чего отсортировать их по значениям ответов. Может сводиться к задачам классификации или регрессии. Часто применяется в информационном поиске и анализе текстов.

Задача прогнозирования (forecasting) отличается тем, что объектами являются отрезки временных рядов, обрывающиеся в тот момент, когда требуется сделать прогноз на будущее. Для решения задач прогнозирования часто удаётся приспособить методы регрессии или классификации, причём во втором случае речь идёт скорее о задачах принятия решений.


Web Mining






Рекомендательные системы

Одно из наиболее популярных приложений интеллектуального анализа данных и машинного обучения в сфере интернет-бизнеса. Данная система анализирует поведение пользователей интернет-сервиса, после чего может давать оценку предпочтения пользователям того или иного объекта рекомендаций. Объектами рекомендаций могут служить товары в интернет-магазине, набор разделов веб-сайта, медиа-контент, другие пользователи веб-сервиса.




Проектирование рекомендательной системы

Python — высокоуровневый язык программирования общего назначения, ориентированный на повышение производительности и читаемости кода []. Python стал общепринятым языком для многих сфер применения науки о данных. Он сочетает в себе мощность языков программирования и простоту использования предметно-ориентированных скриптовых языков типа MATLAB или R. В Python есть библиотеки для загрузки данных, визуализации, статистических вычислений, обработки естественного языка, обработки изображений и многого другого.



Проектирование рекомендательной системы

Django — фреймворк для создания приложений разного уровня сложности. Одним из его основных преимуществ является то, что необходимо продумать только логику приложения. Один из самых популярных фреймворков благодаря встроенной панели администратора, генератору моделей и гибкому ORM. К его отличительным особенностям также относится разделение контента и представления с помощью шаблонов и абстрагирование от низкого уровня баз данных. Может быть использован для создания практически чего угодно — от систем управления контентом и вики, до социальных сетей и новостных сайтов.



Проектирование рекомендательной системы

* здесь будет описание моей системы *



Демонстрация приложения

* здесь будет переход к live demo *