

Исследование различных методов построения рекомендательных систем для фильмографии на основе анализа эмоциональной окраски отзывов и оценок пользователей.

Козлов Е.А. (федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Научный руководитель – кандидат технических наук Штенников Д.Г.
(федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»)

Аннотация:

В докладе рассматриваются различные подходы в построении рекомендательных систем для фильмографии. С использованием открытых данных, содержащих отзывы пользователей и их оценок. Данная статья посвящена сравнительному исследованию алгоритмов при создании рекомендательной системы для ИС.

Введение. В современном мире онлайн сервисов, персонализированное потребление контента является преимуществом перед конкурентами. В статье, на примере сервиса онлайн-кинотеатра, рассмотрены различные модели и алгоритмы: рекомендательная система с использованием сингулярного разложения матрицы, метода ближайших соседей, альтернативного метода наименьших квадратов, и методов анализа контента.

Основная часть.

В 2006 году компания Netflix объявила конкурс на создание крупнейшей рекомендательной системы для фильмографии. На основе большого объема данных были созданы и сравнены миллионы моделей. Соревнование продлилось 4 года, в результате которого была создана ИС с использованием рекомендательного сервиса. В 2006 году Netflix был средним бизнесом по аренде DVD. На данный момент Netflix крупнейший онлайн кинотеатр в мире, с собственными киностудиями и производством, глупо отрицать эффект современных методов анализа данных в этом изменении.

Рекомендательные системы персонализируют выдачу контента под пользователя, они повышают эффективность ИС и, как следствие, прибыль. В статье, сравниваются 5 основных подходов: : рекомендательная система с использованием сингулярного разложения матрицы, метода ближайших соседей, альтернативного метода наименьших квадратов, и методов анализа контента и нейронная сеть для решения задачи ранжирования.

при построении РС на одном наборе данных, с содержанием естественного языка, для обработки которого используются рекуррентные нейронные сети и предобученные методы векторизации.

Выводы.

В результате работы модели сравниваются на различных метриках качества. От числовых, аналитических: MAP@k, nDCG@k до покрывающих метрик, которые показывают насколько хорошо модели работают с разнообразием контента. На основе этих метрик, можно выбрать оптимальное(ые) решение(я) для определённой задачи.

Козлов Е.А. (автор)

Подпись

Штенников Д.Г. (научный руководитель)

Подпись