Адмиралтейский район Санкт-Петербурга

ГБОУ СОШ № 255

С углубленным изучением предметов художественно-эстетического цикла

Рекомендательные системы

Выполнили:

Антипова Е., Геращенко С.,

Камилова Н., Сердукова П., учащиеся 10 класса А

Руководитель работы:

Ярмолинская М.В.

Санкт-Петербург

2024

Оглавление

[Введение 1](#_Toc163514507)

[Цель проекта 1](#_Toc163514508)

[Задача рекомендательной системы 1](#_Toc163514509)

[Актуальность 1](#_Toc163514510)

[Основа 2](#_Toc163514511)

[Рекомендательные системы 2](#_Toc163514512)

[Зачем пользователям рекомендательные системы? 2](#_Toc163514513)

[Зачем сервисам рекомендательные системы? 2](#_Toc163514514)

[Библиотеки 2](#_Toc163514515)

[Виды рекомендательных систем: 4](#_Toc163514516)

[Коллаборативная фильтрация. Метрический метод 4](#_Toc163514517)

[Применение метода 5](#_Toc163514518)

[Формулы для нахождения близости 5](#_Toc163514519)

[Формулы для выставления оценок на основе метрики близости 5](#_Toc163514520)

[Алгоритм 6](#_Toc163514521)

[Коллаборативная фильтрация. Метод на основе разложения матрицы 6](#_Toc163514522)

[Алгоритм 6](#_Toc163514523)

[Виды метрик качества рекомендаций рекомендательных систем 7](#_Toc163514524)

[\*\*Естественный язык 8](#_Toc163514525)

[Обработка естественного языка (NLP) 9](#_Toc163514526)

[Варианты использования MLP 9](#_Toc163514527)

[Источники 10](#_Toc163514528)

# Введение

## Цель проекта

Ознакомление с рекомендательными системами и их составляющими, применениями. Формирование общего мнения технологии.

## Задачи

1. Собрать общую информацию по теме рекомендательных систем.
2. Проанализировать алгоритмы рекомендательных систем.
3. Найти первоисточники в интернете с документацией по библиотекам.
4. Структурировать информацию по теме.
5. Создать общий документ с описанием алгоритмов и библиотек, используемых для создания рекомендательных систем.

## Задача рекомендательной системы

Задача рекомендательной системы – проинформировать пользователя о товаре, который ему может быть наиболее интересен в данный момент времени. Клиент получает информацию, а сервис зарабатывает на предоставлении качественных услуг. Услуги — это не обязательно прямые продажи предлагаемого товара. Сервис также может зарабатывать на комиссионных или просто увеличивать лояльность пользователей, которая потом выливается в рекламные и иные доходы.

## Актуальность

Одним из бурно развивающихся направлений совершенствования индустрии электронной коммерции является развертывание рекомендательных систем - инструментов автоматической генерации предложений по услугам на основе изучения персональных потребностей клиентов. Они позволяют улучшить пользовательский опыт, предоставляя персонализированные рекомендации по товарам, услугам, контенту и многому другому.

Актуальность рекомендательных систем заключается в их способности повысить удовлетворенность пользователей, увеличить продажи и конверсию, снизить отток пользователей и улучшить общую пользовательскую реакцию на предоставляемый контент. С развитием технологий машинного обучения и искусственного интеллекта, рекомендательные системы становятся все более точными и эффективными. Они способны адаптироваться к изменяющимся предпочтениям пользователей и учитывать их поведенческие данные для предоставления наиболее подходящих рекомендаций.

Мировой рынок рекомендательных технологий в 2020 г. был оценен в 1,77 млрд долл. США, прогнозируемые темпы роста на период с 2021 по 2028 гг. составляют около 33% ежегодно. К 2028 г. объем рынка должен достигнуть значения в 17,3 млрд долл. США. Наиболее крупным рынком для рекомендательных систем является регион Северной Америки, его доля от общемирового составляет 33%. В период 2021‒2028 гг. прогнозируется ежегодный рост на 34,4% гибридных рекомендательных систем, используемых, например, Netflix.

# Основа

## Рекомендательные системы

**Рекомендательные системы** – это системы фильтрации информации, предоставляющие пользователям персонализированную рекомендацию по объектам на основе хранимых сервисом данных о пользователях и объектах.

### Зачем пользователям рекомендательные системы?

Рекомендательные системы обеспечивают для пользователя отвечающий его потребностям контент. Обеспечивают экономию времени на поиск этого контента, а также обеспечивают снижение интенсивности коммуникации с клиентом.

### Зачем сервисам рекомендательные системы?

Для самого сервиса внедрение рекомендательных систем – это увеличение некоторых целевых метрик сервиса. В целом метрик удовлетворенности пользователя. Это экономия на коммуникациях, т.е. не нужно тратить деньги на какие-то звонки, сообщения и пуш-уведомления и т.д. А также дополнительный канал продаж для своих услуг.

## Библиотеки

Существует множество библиотек, которые предоставляют инструменты для создания и реализации рекомендательных систем. Все необходимые и самые часто используемые инструменты разбросаны по разным библиотекам. Более того, каждая из таких библиотек имеет много уникальных особенностей, к которым нужно приноровиться (например, разные форматы данных на вход).

«Классические» подходы для работы с рекомендательными системами — библиотеки Implicit и LightFM.

Implicit реализует базовые методы для работы с неявным фидбэком. То есть для создания рекомендательных систем используются не только ответы самого пользователя, например, какой он рейтинг поставил, но и неявные действия: человек долго задержался в ленте на объекте или просто на него кликнул.

LightFM помогает сделать систему, которая будет учитывать не только взаимодействия пользователей и объектов, но и их фичи — атрибуты. Для пользователей это пол, возраст, доход, регион и так далее. Для объектов, например, фильмов — это жанр, режиссёр, бюджет, аннотация фильма.

RecTools - библиотека, где собраны самые часто используемые модели для рекомендательных систем. Также с её помощью можно максимально просто и быстро оценивать необходимые метрики.

Другая модель в RecTools — популярное, она самая базовая. Она ищет в данных, с чем пользователи чаще всего взаимодействуют — что чаще всего смотрят или покупают. Это легко сделать и проинтерпретировать, но опыт не самый лучший. Потому что в топе будут попадаться либо фильмы, о которых пользователь и так слышал, либо они не будут подходить ему по интересам. Поэтому в RecTools подключены и другие библиотеки.

Последняя модель — Pure SVD, которая реализует векторизацию пользователей и объектов. Она берёт исходную разреженную матрицу взаимодействий и представляет её в виде произведений трёх матриц. В этом суть SVD — Singular Value Decomposition. Из этих трёх матриц складываются две матрицы, которые потом интерпретируются как матрица пользователей и матрица объектов. Там каждая строчка — это вектор пользователя (объекта).

Surprise - простая библиотека Python для построения и анализа рекомендательных систем. Она предоставляет различные алгоритмы совместной фильтрации, такие как SVD, KNN и другие.

Это лишь несколько примеров библиотек, которые могут быть использованы при создании рекомендательных систем. Каждая из них имеет свои особенности и подходит для определенных задач и типов данных. Выбор конкретной библиотеки зависит от конкретных требований проекта и уровня экспертизы в области машинного обучения.

## Виды рекомендательных систем:

1. Методы коллаборативной фильтрации – используют только данные о взаимодействии пользователей и объектов
* Метрические методы
* Методы на основе матричных разложений
1. Методы на основе контента – используют данные о пользователях и объектах
2. Гибридные рекомендательные системы – используют данные о пользователях, объектах и их взаимодействиях

## Коллаборативная фильтрация. Метрический метод

**Коллаборативная фильтрация** — это технология, которая позволяет прогнозировать предпочтения конкретного пользователя интернет-ресурса, сравнивая его интересы с интересами других посетителей ресурса. Основное допущение метода состоит в том, что те, кто одинаково оценивал какие-либо предметы в прошлом, склонны давать похожие оценки другим предметам и в будущем. На основании этой информации в рекомендательном разделе предлагаются те товары, которыми интересовалась аудитория сайта, а конкретный пользователь — еще нет. Этот метод обработки интересов пользователей далеко не новый.С помощью коллаборативной фильтрации можно, например, прогнозировать, какая музыка понравится пользователю, имея неполный список его предпочтений (симпатий и антипатий).

Прогнозы составляются индивидуально для каждого пользователя, хотя используемая информация собрана от многих участников.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | Бег | Идиот | Мимино | Покровские ворота | Римские каникулы |
| 1 | 3 |  |  | 7 | 9 |
| 2 | 8 | 9 | 6 | 4 | 6 |
| 3 | 9 | 8 | 4 |  | 5 |
| 4 |  |  | 5 | 7 | 7 |
| 5 | 8 | 8 |  | 3 |  |

### Применение метода

1) Ищем для определенного пользователя ближайших соседей, то есть тех, кто поставил максимально похожие оценки (пользователь 5 поставил высокие оценки фильмам «Бег» и «Идиот», пользователи 2 и 3 тоже посмотрели эти фильмы и поставили им высокие оценки. Значит пользователи 2 и 3 соседи пользователя 5)

2) Смотрим какие фильмы, из просмотренных соседями, не смотрел наш пользователь (Пользователь 5 не смотрел фильм «Мимино», пользователи 2 и 3 посмотрели и поставили оценки 4 и 6. Ищем усредненную оценку, в нашем случае будет 5)

3) Рекомендуем эти фильмы пользователю (Выставляем усредненную оценку 5 пользователю на основе его соседей)

### Формулы для нахождения близости

- Обозначения:

Пользователь – u, Объект – i, Настоящая оценка, которую выставил пользователь – r, Предсказание - $\hat{r}$.

- Каждый пользователь характеризуется вектором оценок (строка в таблице)

- Чтобы определить меру близости этих векторов нужно использовать косинус угла между векторами

Пример: Вектор пользователя 5 состоит из [8, 8, - , 3, - ] $\vec{u\_{5}}=[8, 8, nan, 3, nan]$

 Вектор пользователя 3 состоит из [9, 8, 4, - , 5] $\vec{u\_{3}}=[9, 8, 4, nan, 5]$

ц- Берем только те значения, которые есть у обоих пользователей

(берем 1 и 2 координату)

- Формула косинуса между векторами:

$$\cos(\left(u\_{5}, u\_{3}\right)= \frac{\vec{u\_{5}}∙\vec{u\_{3}}}{\left|\vec{u\_{5}}\right|\left|\vec{u\_{3}}\right|})=\frac{8∙9+8∙8}{\sqrt{8^{2}+8^{2}}\sqrt{9^{2}+8^{2}}}=0.998$$

### Формулы для выставления оценок на основе метрики близости

Отбираем два ближайших соседа и усредняем их оценки, выставленные фильму «Мимино», чтобы выставить оценку для пользователя 5

Формулы:

$$\hat{r\_{ui}}=\frac{∑\_{j\in N(u)}r\_{ji}}{\left|N(u)\right|}$$

$$\hat{r}5,Минимо=\frac{∑\_{j\in N[2,3]}r\_{j,Минимо }}{\left|\left\{2, 3\right\}\right|}=\frac{6+4}{2}=5$$

### Алгоритм

1. Выбрать метрику близости, произвести ее расчет
2. Выбрать n ближайших соседей для пользователя
3. Рассчитать оценки объектов для пользователя на основе известных оценок соседей
4. Отсортировать объекты по убыванию оценок и показать пользователю самый высокорейтинговый

#### Плюсы и минусы

|  |  |
| --- | --- |
|  Плюсы |  Минусы |
| Простота реализации | Нужны оценки пользователей |
| Интерпретируемость | Невозможность предсказания для нового пользователя |
| Низкая вычислительная сложность | Невозможность предсказания для нового объекта |
| Легкость добавления новых объектов и пользователей |  |

## Коллаборативная фильтрация. Метод на основе разложения матрицы

### Алгоритм

Мнимую оценку (т.е. модельное предположение) объекта пользователем можно представить в следующем виде:

 $\hat{r}=μ+b\_{u}+b\_{i}$ $μ-средняя оценка объектов$

 $b\_{u}-смещение пользователя$

 $b\_{i}- смещение объекта$

Представим, что мы уже получили оценку объекта пользователем, тогда оценить ошибку нашего предсказания можно с помощью функции:

$$L\left(u, i\right)=(r\_{ui}-\hat{r}\_{ui})^{2}$$

Научившись измерять ошибку предсказания, мы должны по данным таблицы взаимодействия подобрать такие коэффициенты смещения пользователя и смещения объекта, чтобы минимизировать ошибку предсказания оценки

$$\sum\_{(i, u)}^{}\left(r\_{ui}-μ-b\_{u}-b\_{i}\right)^{2}\rightarrow \begin{array}{c}mⅈn\\b\_{u}b\_{i}\end{array}$$

∑ - сумма предсказаний по всем фильмам для каждого пользователя

#### Более полные формулы с использованием векторов

##### Модель

 $r\_{ui}=μ+b\_{u}+b\_{i}+q\_{u}^{T}q\_{i}$ $μ-средняя оценка объектов$

 $b\_{u}-смещение пользователя$

 $b\_{i}- смещение объекта$

 $q\_{u}-вектор пользователя$

 $q\_{i}-вектор объекта$

По данным таблицы взаимодействия подберем такие коэффициенты смещения пользователя, смещения объекта и вектора пользователя, вектора объекта, индивидуальные для каждого пользователя и фильма, чтобы минимизировать ошибку предсказания оценки

$$\sum\_{\left(i, u\right)}^{}\left(r\_{ui}-μ-b\_{u}-b\_{i}-q\_{u}^{T}q\_{i}\right)^{2}\rightarrow b\min\_{b\_{u}, b\_{i},q\_{u},q\_{i}}$$

### Виды метрик качества рекомендаций рекомендательных систем

Существует множество метрик качества рекомендательных систем. Вот некоторые из них:

На уровне машинного обучения используются следующие метрики:

* MAE (Mean Absolute Error) — средняя абсолютная ошибка;
* RMSE (Root Mean Squared Error) — среднеквадратичная ошибка.

На уровне поиска информации оцениваются следующие метрики:

* Accuracy — общий показатель точности для всех классов;
* Precision — отношение соответствующих рекомендаций к длине всего списка представленных рекомендаций;
* Recall — полнота;
* F-мера — компромисс между точностью и отзывом;
* ROC — соотношение ложных и истинных срабатываний.

На уровне взаимодействия пользователя и компьютера оцениваются следующие метрики:

* Diversity — разнообразие рекомендуемых объектов;
* Trust — пользовательское доверие рекомендательной системе;
* Novelty — новизна рекомендуемых объектов;
* Serendipity — случайность рекомендуемых товаров.

На уровне разработки программного обеспечения оцениваются следующие метрики:

* Real-time — производительность рекомендательных систем в реальном времени;
* Robustness — прочность или надежность рекомендательной системы;
* Scalability — масштабируемость рекомендательной системы.18:54

# \*\*Естественный язык

Естественный язык — в лингвистике и философии языка язык, используемый для общения людей (в отличие от формальных языков и других типов знаковых систем, также называемых языками в семиотике), то есть человеческий язык, который, в отличие от искусственных языков, не создан целенаправленно, искусственно, а исторически сложился в процессе развития человеческого общества. Насчитывается до 7 тыс. естественных языков.

Словарь и грамматические правила естественных языков определяются практикой применения и не всегда бывают формально зафиксированы.

## Обработка естественного языка (NLP)

Технология машинного обучения, которая дает компьютерам возможность интерпретировать, манипулировать и понимать человеческий язык, называют обработкой естественного языка (NLP). Сегодня организации имеют большие объемы голосовых и текстовых данных из различных каналов связи, таких как электронные письма, текстовые сообщения, новостные ленты социальных сетей, видео, аудио и многое другое. Они используют программное обеспечение NLP для автоматической обработки этих данных, анализа намерений или настроений в сообщении и реагирования на человеческое общение в режиме реального времени.

Обработка NLP присутствует в повседневных взаимодействиях со всеми видами машин. Когда ты вводишь вопрос в поисковую систему, NLP анализирует твое намерение, чтобы затем предоставить релевантные результаты. Виртуальные помощники, такие как умные колонки или чат-боты, полагаются на обработку естественного языка для взаимодействия с людьми. Другие применения NLP – это автоматически генерируемые переводы и подписи, сортировка сообщений, проверка орфографии и грамматики, распознавание рукописного или напечатанного текста и преобразование текста в речь.

Обработка естественного языка использует машинное обучение, чтобы научить компьютеры понимать и переводить человеческий язык. Чем больше они узнают, тем лучше они могут понимать текст в устной или письменной форме, классифицировать или переупорядочивать его, переводить и взаимодействовать с ним. Для распознавания и обработки естественного языка используются такие библиотеки как NLTK, spaCy, Gensim, Natasha.

## Варианты использования MLP

Анализ эмоций: NLP может классифицировать эмоции в тексте как положительные, отрицательные или нейтральные. Facebook делает это с пользовательским контентом, но бренды также используют этот метод, чтобы понять, что клиенты думают об их продуктах.

Извлечение текста: NLP может находить соответствующие термины в тексте любого размера и извлекать или обрабатывать их.

Классификация тем: текст можно разделить на отдельные части по разным темам:

Обработка документов: это позволяет пользователям, не знакомым с программированием или обучением искусственного интеллекта, сообщать компьютеру, что делать со стопкой цифровых или виртуальных документов, например, обрабатывать формы или рассчитывать затраты, возвраты и т. д.

Генерация текста: хотя искусство, созданное с помощью искусственного интеллекта, все еще сомнительного качества, обработка естественного языка может генерировать разборчивый и значимый текст, например, сводку спортивных результатов. При достаточно большом размере выборки NLP может имитировать стиль конкретного автора и соответственно переписывать текст.

Машинный перевод. Сайты-переводчики Google Translate, «Яндекс.Переводчик» и другие делают перевод текстов на другие языки доступным и быстрым.

# Источники

<https://www.dissercat.com/content/issledovanie-i-razrabotka-algoritmov-rekomendatelnykh-sistem-na-osnove-grafovykh-modelei-dan>

<https://rdc.grfc.ru/2023/05/recommendation_services_methods_and_regulation/>

<https://habr.com/ru/articles/773126/>

<https://tproger.ru/articles/kak-ustroena-rectools-opensource-biblioteka-dlja-rekomendatelnyh-sistem>