

## Интро

На сегодняшний день сложно себе представить нашу жизнь без рекомендательных систем которые уже прочно проникли в нашу жизнь. Они не только рекомендуют ещё одно видео с котиками, но и формируют музыкальные листы, новостные ленты, похожие товары или фильмы, оптимальные маршруты, друзей по интересам и даже репозитории на гитхабе!

## Постановка задачи

Участникам необходимо используя имеющиеся открытые данные популярной социальной сети построить рекомендательную систему, которая способна по имеющимся подпискам пользователя предсказать наиболее релевантные сообщества.

## Доступные данные

Участникам доступны следующие источники данных:

- Информация о пользователях
- Информация о подписках пользователя
- Информация о сообществах

## Оценка качества

Для оценки качества рекомендаций используется метрика [MAP](#), где precision для каждого пользователя считается по указанному алгоритму:

```
def precision_at_k(predicted, actual, k=10):  
    predicted = predicted[:k]  
    denominator = min(len(actual), k)  
    intersection = set(predicted) & set(actual)  
    return float(len(intersection)) / denominator
```

Далее результат усредняется по всем пользователям который участвовали в оценке.

Во время подготовки тестового набора данных было семплировано 1000 случайных пользователей, которые были подписаны на более чем 5 сообществ. Далее подписки каждого пользователя были разделены на приватную и публичную часть. Публичная часть отправляется сервером оценки в систему рекомендаций которую строят участники. В качестве ответа сервер оценки ожидает *отсортированный по релевантности* список сообществ для каждого пользователя в публичном наборе данных. Далее на сервере оценки происходит расчёт вышеописанной метрики с использованием скрытых приватных данных. Лучшая полученная оценка будет использоваться для построения лидерборда.

Для оценки качества необходимо отправить запрос на сервер {адрес} который содержит название команды и URL адреса системы рекомендаций. Запрос на оценку можно отсылать не чаще чем раз в 5 минут.

## Базовое решение

Командам доступно базовое решение которое рекомендует топ-10 самых популярных сообществ по количеству подписчиков. Так же базовое решение содержит код который позволяет

обрабатывать входной запрос от сервера оценки качества. Помимо этого, реализован удобный web-интерфейс для взаимодействия с рекомендательным движком.

s3://nonauth-friendsday-kiur1wee/common/data\_science/recommender\_template.zip

## Алгоритмы построения рекомендательной системы

В качестве базового решения мы предлагаем использовать нейросетевой алгоритм который уже реализован в фреймворке [StarSpace](#). Для обучения репрезентаций сообществ необходимо сформировать входной файл в следующем формате:

```
pub_1 pub_2 pub_3 pub_99 pub_98
pub_21 pub_22 pub_23
```

Где каждая строчка – это сообщества пользователя на которые он подписан

pub\_%id% - префикс с индикатором сообщества. В качестве разделителя должен использоваться символ табуляции.

Для запуска обучения необходимо выполнить команду:

```
./starspace train -trainFile 'starspace_input.tsv' -label 'public' -trainMode 1 -dim 64 -model 'starspace_output' -negSearchLimit 300 -maxNegSamples 10
```

Результатом обучения будет векторное представление сообществ в 64-мерном пространстве которое в дальнейшем можно использовать для формирования рекомендаций.

Интересы пользователя можно представить в виде среднего его подписок в обученном пространстве сообществ. Далее необходимо найти наиболее похожие сообщества по cosine similarity.

В конечном итоге необходимо:

- Реализовать функцию рекомендаций:

```
recommender_template.zip
```

```
recommendation_engine.py
```

```
get_recommendations_from_publics
```

- Оценить качество рекомендаций
  - [http://ec2-3-120-132-157.eu-central-1.compute.amazonaws.com:8080/evaluate?team\\_name=%ИМЯ\\_КОМАНДЫ%&endpoint=http://%АДРЕС\\_СЕРВЕРА%:6767/endpoint](http://ec2-3-120-132-157.eu-central-1.compute.amazonaws.com:8080/evaluate?team_name=%ИМЯ_КОМАНДЫ%&endpoint=http://%АДРЕС_СЕРВЕРА%:6767/endpoint)
- Подняться как можно выше на Leaderboard
  - <http://ec2-3-120-132-157.eu-central-1.compute.amazonaws.com:8080>