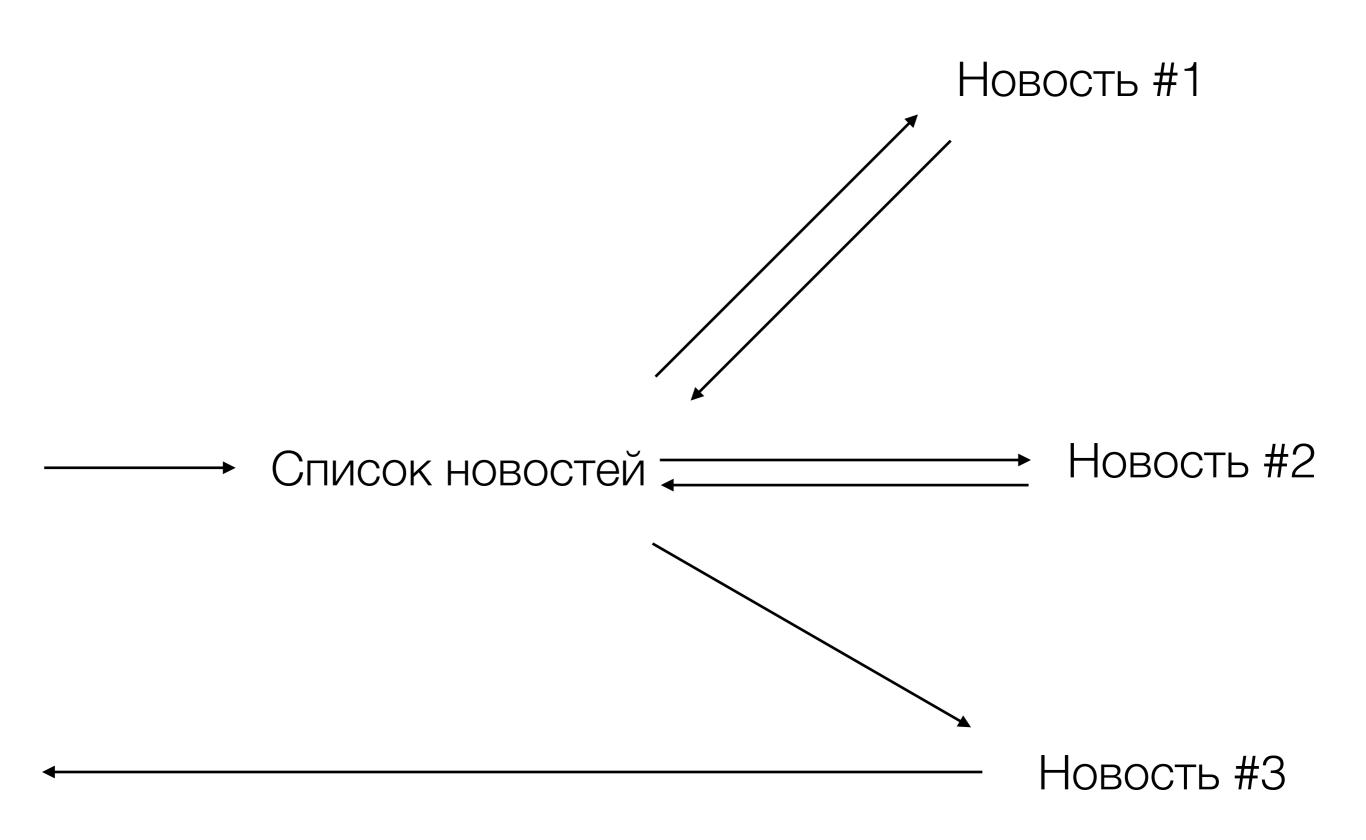
tvrain.ru recommender

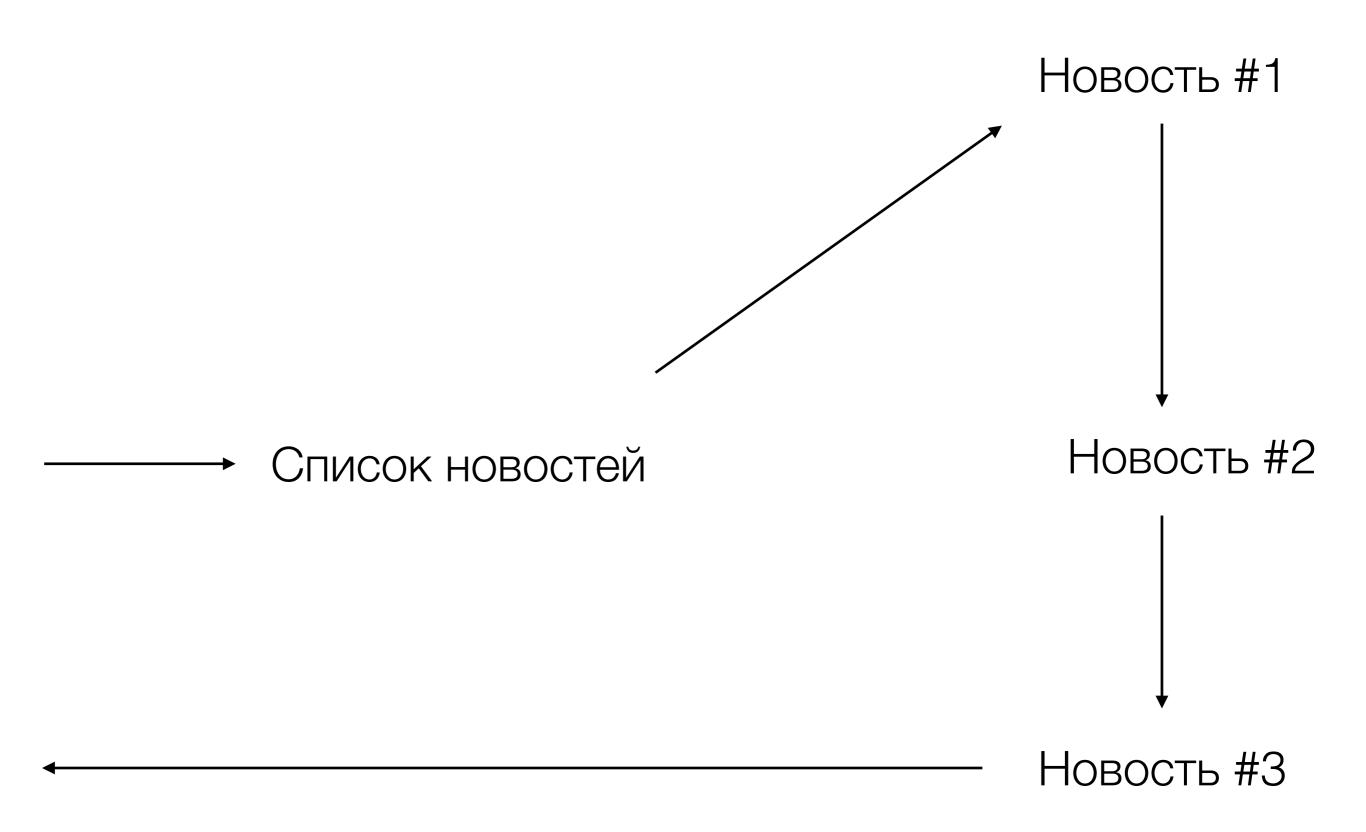
Рекомендательная система для телеканала «Дождь»



Что за рекомендательная система?

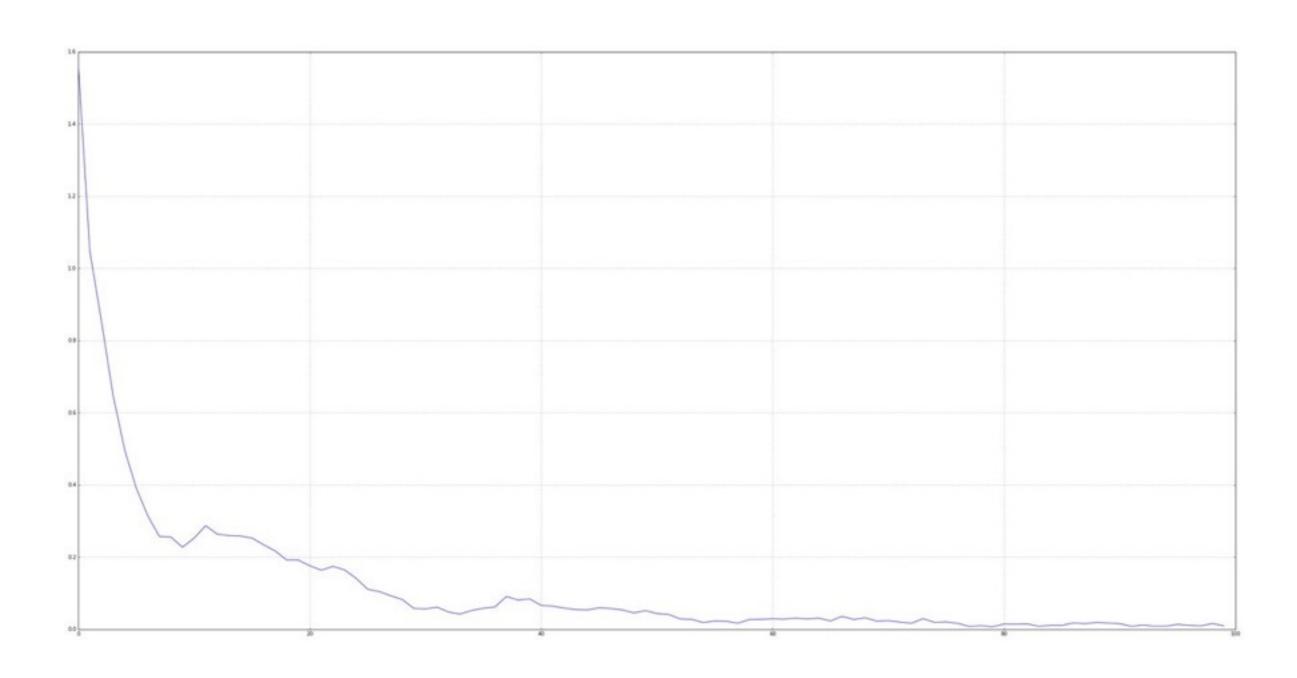


Как это было



Как должно быть

Новости — это не фильмы



Что у нас получилось

News recommender.

- Советник президента по интернету допустил блокировку Google и Facebook
- Герман Клименко, глава Liveinternet: надеюсь, «закон Лугового» постигнет участь закона о блогерах – применять будут в крайних случаях и на индустрию не повлияет
- Герман Клименко: «Я только советник».
- Гитлер, Микки Маус и быстрые девочки на Олимпиаде
- Самойлов подал в суд на Троицкого
- Эбонитовая палочка и баночка «Ягуара»: музей энергетики в Ганновере
- Богатейшие люди мира потеряли более \$194 млрд в первые дни 2016 года
- Гетто иностранных агентов.
- Мэр Львова Андрей Садовый о праздновании Дня Победы: «Мы молимся, чтобы никогда больше такая беда на нашу землю не пришла»
- Проверка паспортов и нехватка мечетей. Подготовка к Курбан-Байраму на фоне стычки в Москве

https://tvrain.ru/news/google i facebook-402534/

https://tvrain.ru/teleshow/muzejnyj_smotritel/ebonitovaja_palochka_i_bar

https://tvrain.ru/teleshow/vechernee_shou/german_klimenko_ja_tolko_sc

Submit

Рекомендации

- Кто будет советовать Путину, что делать с интернетом.
- Герман Клименко: «Телеграм или будет сотрудничать, или будет закрыт».
- Вена. Музей контрацепции

И всё?

```
ndels. features. cosine import cosine_similarity_features
                                                                                                                                                           second_set_probability = first_set_len / all_items_num

value = {

all_items_num

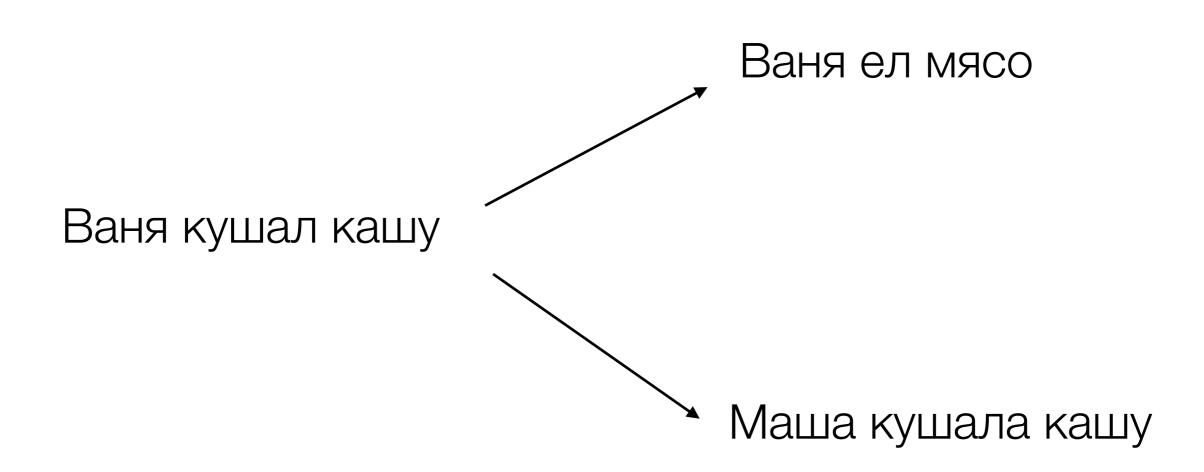
all_items_num
                                                                                                                                                             first_set_probability = first_set_set_lon_num / all_items_num
                                                                                                                                                                   rst_set_non_probability = (intersection_num / all_items_num) *** 2
                                                                                                                                                                                                      -n: " of first " for example, sum of views of first vio
                                                                                                                                                          Cosine similarity value = (
                                                                                                                                                              intersection_probability /

* second_set_probability)
. RandomForestClassifier()
                                                                                                                 data[-1]))
generate_features(data, val=None):
     features.extend(topics_similarity_features(data)

features.extend(cosine_similarity_features)
                                                                                                            crawlers
                                                                                                                   ime_crawler.py
                                                                                                      ▼  ata
  features = []
    for raw in data:
                                                                                                                   🔑 __init__.py
                                                                                                                                                                                                          rt_articles, output_article):
                                                                                                                   data_utils.py
                                                                                                                                                                                                          cles and possible recommended article
                                                                                                       ▼ • models
          if val is None:
                   return features
                                                                                                             features
                      return features, val
                                                                                                                          🔑 __init__.py
                                                                                                                         cosine.py
                                                                                                                          dopics.py
           def generate_data(data):
                                                                                                                   _init__.py
                                                                                                                                                                                                           cle['views'])
                                                                                                                   hybrid_model.py
                           print('Start generate features.')
                     x_true = []
                                                                                                                                                                                                           (cle['views'])
                             for urls in data. sequences. fund():
                                                                                                                   nmf_model.py
                       y_true = []
                                      features = generate_features(da
                         X_false = []
                                                                                                                                                                                                        le_views.intersection(input_article_views)), 's id's
                                                                                                            scripts
Juind Jurity
                          y_false = []
                                        if len(features[0]) = 163:
                                                                                                                   count_all_users_views.py
                                                  x true. append (features [0])
 recommen
                                                  y_true.append(features[1])
                                                                                                                   graph_by_url.ipynb
                                                                                                                    load_logs_to_mongodb.py
If the of
ist popular ar
                                                                                                                    load_sequences_to_mongodb.py
                                                      print(features, urls)
load_topics_to_mongodb.py
                                         print('Start generate random
len(x_true) != len(x_true) != generate_features = generate_features [0]) != features = generate_features [0]) != if len(features [0]) != len(features [0]) !
                                                                                                             templates |
                                                                                                             gitignore.
                                                                                                             app.py
                                                                                                                                                                                                           mp.strip('GET').strip('Z')
                                                                                                             predict.py
                                                                                                             README.md
                                                                                                                                                                                                          imestamp.split('.')[:-1])
                                                                                                             requirements.txt
                                                                                                                                                                                                          .strptime(timestamp, '%Y-%m-%dT%H:%M:%S')
                                                                                                             ii scrot.png
                      tine we need to fit model
                                                                                                                                                                        or not user:
                                                                                                                                                            Symme irl.startswith('https://tvrain.ru/'):
                                                                                                                                             le_urls.add(url)
                                                                                                                                                          Atinue
                                                                                                                                               cle_views[url].append(user)
                                                                                                                  ign, find one (f'
                                                          tictle articles
                                                                                                                                            Save time of only first our
                                                                                                                                           url not in article
                                                                               ral num
```

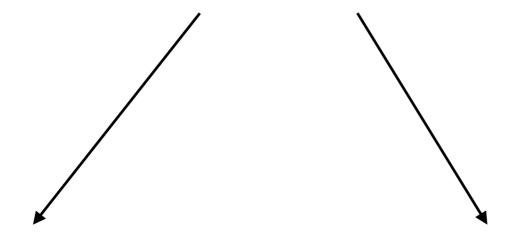
А как рекомендовать?

Контентная модель



Topic Modelling

Ваня кушал кашу



Тема «Поедание каши» - 22%

Тема «Ваня» - 15%

Коллаборативная фильтрация

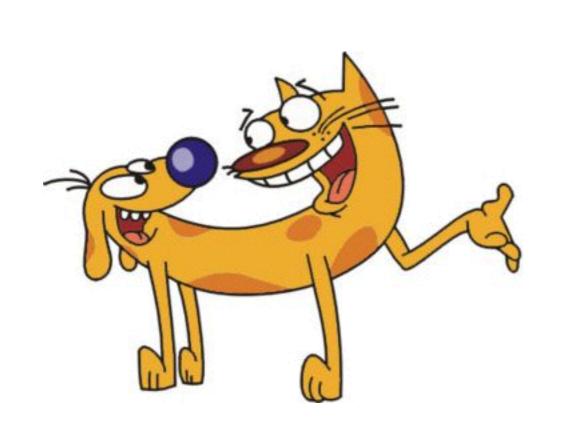
	Путин подписал	Савченко отпустили	Астахов заявил
Творог	+	+	+
Сева		+	
Дима			
Олег	+	+	

Считаем похожесть новостей

$$K = \frac{|A \cap B|^2}{|A| \cdot |B|}$$

Находим новости с максимально похожими посетителями, предлагаем их

Гибридная модель



Получаем результаты контентной модели, поучаем результаты коллаборативной фильтрации

и все кидаем в Random Forest

Гибридная модель

Новость, которую юзер уже прочитал, похожа на новость про Путина, а новость, которую юзер может прочитать похожа на новость про Савченко.

Кроме того, пересечение посетителей новостей - 5%

Мне кажется, эта новость подойдет этому юзеру с вероятностью 20%

У нас всё.

github.com/tvorozid/recommendation_system xenx.ml:8080