

Проект:
Прогнозирование хэштегов по
фотографиям в профиле социальной сети

Команда:
В. Савкина,
П. Петрова,
З. Мазунина

<https://github.com/zozunin/tag4pic-project>

Задачи

- Обновить корпус фотографии Harrison (<https://github.com/minstone/HARRISON-Dataset>)
 - Выделить наиболее популярные хэштеги с помощью специализированных мониторинговых сайтов
 - Скачать новые данные с помощью библиотеки instaloader (<https://instaloader.github.io/>)

→ Общий корпус фотографий и хэштегов составил 80K постов

- Выделить наиболее перспективные модели, адаптировать под новые данные, обучить и проанализировать результаты
 - ALS recommender + Deep Neural Network
 - Word2Vec+Resnet
- Создать tg-бота

Самые популярные хэштеги

- instagood
- love
- fashion
- photooftheday
- art
- photography
- beautiful
- picoftheday
- nature
- happy
- cute
- travel
- style
- instadaily
- summer
- beauty
- fitness
- food
- instalike
- friends
- family
- life
- music
- lifestyle
- design
- motivation
- explore
- nofilter
- foodporn
- instamood
- artist
- wedding
- bestoftheday
- workout
- study
- naturephotography
- nails
- tattoo
- landscape
- blackandwhite
- work
- architecture
- car
- cat
- dog
- animals
- travelphotography
- coffee
- adventure
- vscocam

<https://blog.hubspot.com/marketing/instagram-hashtags>

ALS recommender + DNN

- - MobileNetV2 (pre-trained) — эмбединги изображений (deep features)
 - Поиск k похожих изображений: max cosine similarity
 - Поиск хэштегов — рекомендательная система ALS
 - Базовая идея: если множество изображений имеет k одинаковых хэштегов и m уникальных, каждый из m уникальных хэштегов может быть применен к k изображениям
 - $Q \sim U \times P$, U — user(photo) matrix, P — product (hashtag) matrix
 - $Q(18220 \times 1233)$, 73106 stored 1s
 - $\text{user_features}(\text{test_photo}) = \text{mean}(\text{user features of } k \text{ nearest neighbors})$

Updates since 26.05:

- Dimension reduction (PCA): $1280 \times 1 \Rightarrow 206 \times 1$ (кумулятивный вклад $\geq 85\%$)
- Average hashing для создания ключей изображений (исключает наличие одинаковых изображений)
- Добавлена возможность расширения базы фотографий и хэштегов пользователем:
 - embeddings / user (photo) features — аналогично тренировочному сету
 - product features — решение СЛЮ

Resnet+Word2Vec

Модуль 1.

Обучение w2v на базе

последовательностей хэштегов,
соответствующих фотографиям из
корпуса.

- embedding_dim=500 -> 150
- skip_window=3
- Общее количество хэштегов = 1300

Модуль 2.

2 блока обучения

1. Извлечение эмбедингов изображений с помощью Resnet50
2. Двухслойная нейронная сеть для маппинга изображения и хэштега (2 Linear layers)

Модуль 3.

- Поиск top-10 ближайших хэштегов для предсказанного вектора, для измерения расстояния - cosine dist

Resnet+Word2Vec

- Обучение на 30 эпохах (~12 часов)
- Обучение w2v - 100 эпох
- Эксперименты:
 - skip_window = 2 - Recall 24.91 (27 ep)
 - **skip_window = 3 - Recall 25.68 (23 ep)**
 - skip_window = 5 - Recall 25.07 (28 ep)
- hashtag embedding dim = 500 ->150

Resnet+Word2Vec

Источники:

1. HADES: HAShtag Recommendation via DEep Learning for Images in Social Networks - 2020, URL: https://github.com/kuanting/image_hashtag/ - CNN for images+W2V
2. An Efficient and Resource-Aware Hashtag Recommendation Using Deep Neural Networks - 2019, URL: https://www.researchgate.net/publication/332199478_An_Efficient_and_Resource-Aware_Hashtag_Recommendation_Using_Deep_Neural_Networks - Resnet50+1000classes classification+W2V for tags
3. From Fully Supervised to Zero Shot Settings for Twitter Hashtag Recommendation - 2019, URL: <https://arxiv.org/pdf/1906.04914.pdf> - CNN for tweets + W2V for tags
4. Using Tweets Embeddings For Hashtag Recommendation in Twitter - 2018, URL: <https://www.sciencedirect.com/science/article/pii/S1877050918301030> - W2V for tweets+clustering

Сравнение качества

	ALS recommender + DNN	Resnet+Word2Vec
Recall	21.90%=> 24.50%	21.24%=> 25.68%

Лучшие результаты на данных типа Harrison других разработчиков

	Harrison Baseline (VGG-16) https://arxiv.org/abs/1605.05054	VGG-model https://trepo.tuni.fi/bitstream/handle/10024/104479/1540547474.pdf
Recall	21.38%	15%

Original: Animals (not in train/test)



ALS recommender + DNN

#photooftheday, #photography, #cute, #happy, #follow,
#love, #art, #instalike, #instagram, #fashion

Resnet+W2V

#love, #cute, #family, #bestfriend, #happy, #adorable,
#together, #picoftheday, #mylove, #loveyou

Original: Architecture (not in train/test)



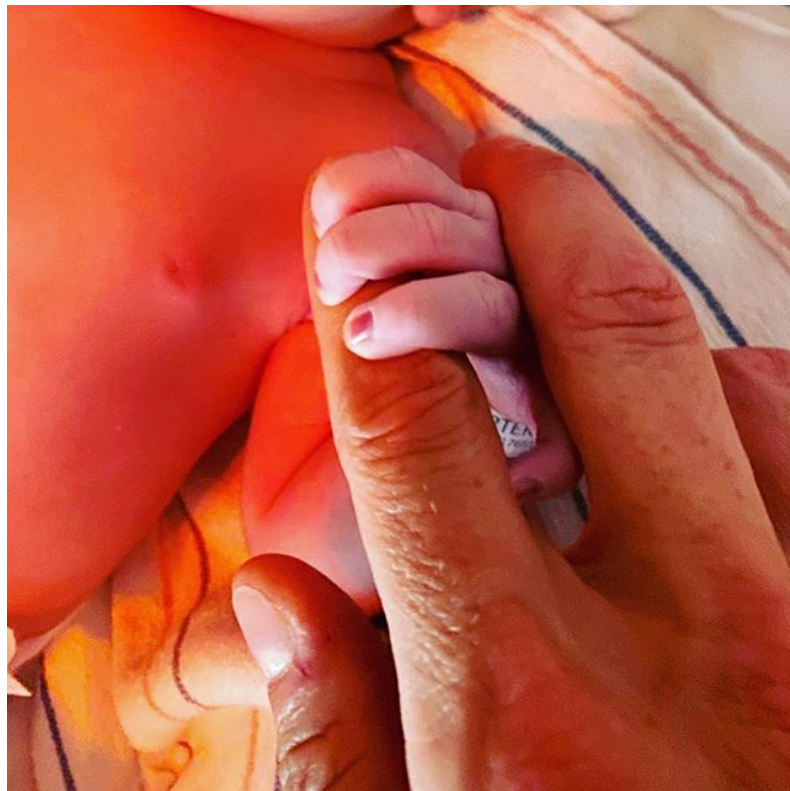
ALS recommender + DNN

#beautiful, #travel, #fitness, #love, #family, #follow,
#happy, #instagood, #gym, #fun

Resnet+W2V

#motivation, #love, #travel, #beautiful, #nature,
#photooftheday, #happy, #study, #success, #workout

Original: Babies (not in train/test)



ALS recommender + DNN

#nails, #nailart, #nailsofinstagram, #nailsart,
#nailsdesign, #manicure, #nail, #naildesigns, #gelnails,
#nailsoftheday

Resnet+W2V

#nails, #nailsofinstagram, #naildesigns, #nailsdesign,
#manicure, #nailsart, #nailpolish, #nailtech,
#nailsoftheday, #gelnails

Original: Cars (not in train/test)



ALS recommender + DNN

#instamood, #explore, #instalike, #follow, #life,
#lifestyle, #design, #followforfollowback, #explorepages,
#bhfyp

Resnet+W2V

#lifestyle, #photooftheday, #travel, #happy,
#photographylovers, #fashionphotography, #style, #likes,
#instamood, #design

Original: Fitness



ALS recommender + DNN

#summer, #follow, #motivation, #lifestyle, #instagram,
#instagood, #travel, #style, #friends, #like

Resnet+W2V

#workout, #bodybuilding, #fitness, #gym,
#fitnessmotivation, #gymmotivation, #gymlife, #crossfit,
#fitnessaddict, #healthylifestyle

Original: Food



ALS recommender + DNN

#foodporn, #art, #life, #photooftheday, #love,
#lifestyle, #friends, #summer, #family, #selfie

W2V+Resnet

#food, #instadaily, #foodie, #foodblogger, #delicious,
#foodlover, #foodphotography, #foodstagram,
#foodgasm, #foodporn

Original: Nature



ALS recommender + DNN

#naturephotography, #lifestyle, #nature, #instagram,
#instagood, #beautiful, #fitness, #life, #instalike, #travel

Resnet+W2V

#naturephotography, #landscapephotography,
#naturelover, #travelphotography, #nature_perfection,
#nature, #naturelovers, #canonphotography,
#flowerphotography, #beautifuldestinations

Original: Selfie



ALS recommender + DNN

#fashion, #love, #girl, #family, #instadaily,
#photooftheday, #style, #bestoftheday, #photography,
#motivation

Resnet+W2V

#nofilter, #beautiful, #photography, #photooftheday,
#likes, #life, #photographer, #picoftheday, #fotografia,
#travelphotography

Original: Travel



ALS recommender + DNN

#cute, #fashion, #style, #love, #photooftheday,
#instagood, #instagram, #beautiful, #happy, #nature

W2V+Resnet

#lifestyle, #photooftheday, #travel, #travelling,
#weddingday, #summer, #stylish, #newcollection,
#traveltheworld, #weddingphotography

Original: Wedding



ALS recommender + DNN

#wedding, #photography, #love, #beauty, #cute, #like,
#photooftheday, #picoftheday, #fashion, #instagood

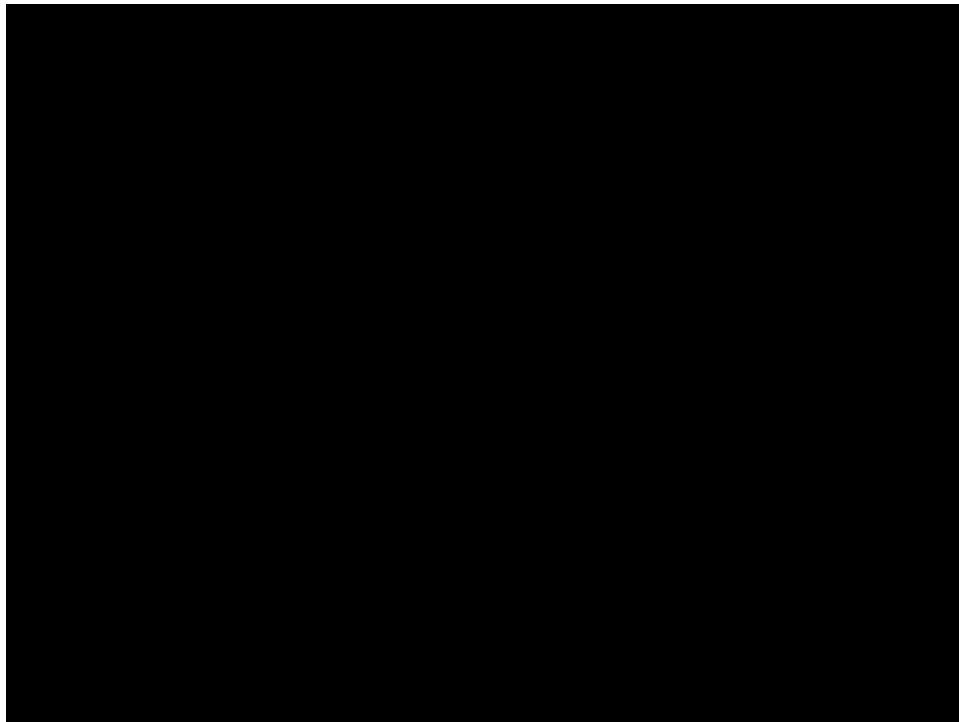
W2V+Resnet

#wedding, #bride, #weddings, #bridgetobe, #groom,
#weddingday, #weddingphotography, #bridal,
#engagement, #weddingplanner

telegram: @tag4pic_bot

- выдает хештеги по присланной картинке;
- возможность выбрать одну из двух представленных моделей;
- возможность посмотреть, какая модель используется;
- возможность добавить хештеги для дообучения 2 модели.

telegram: @tag4pic_bot



Также можно посмотреть демонстрацию здесь: <https://clck.ru/sJNLU>