

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(СПбГУ)

Направление: Прикладные математика и физика



Полуавтоматическое структурирование изображений
в социальной сети с помощью методов машинного
обучения

Санкт - Петербург
2019

Содержание

Введение	3
1 Постановка задачи	4
2 Используемые инструменты	6
3 Сверточные нейронные сети	7
3.1 Введение в теорию сверточных нейросетей	7
3.2 Обзор используемых архитектур	7
4 Подготовка данных	8
4.1 Датасет SUN	8
4.2 База WordNet	8
4.3 Адаптация датасета SUN	8
5 Численные эксперименты	9
5.1 Архитектура программы	9
5.2 Особенности реализации	9
5.3 Полученные результаты	9
5.4 Обсуждение результатов	9
6 Валидация результатов	10
Выводы	11
Список литературы	12

Введение

С каждым днем пользователи социальных сетей создают и потребляют все бóльшие и бóльшие объемы информации, в том числе огромное количество фотографий. Построение системы для быстрой и точной навигации в миллионах изображений — не тривиальная задача. Одно из самых распространенных решений этой проблемы заключается в использовании тегов. Частный случай такого подхода — использование хештегов в социальных сетях.

Хештег — это любое слово или фраза без пробелов, перед которой стоит символ #, который называется *диез* или *решетка*, а в англоязычном варианте — *hash*, отсюда и название. Приведем несколько примеров: *#masterwork*, *#spbu*, *#htaglovesport*. Обычно в браузере или приложении хештег отображается как гипертекст, кликнув по которому можно получить список публикаций, снабженных таким же тегом.

Кроме простоты и удобства теги обладают еще одним полезным свойством — они позволяют не думать об иерархии структурируемой информации. Например, набор изображений можно разложить по папкам, создав иерархию по датам, геолокациям или авторству. Причем в отдельных случаях подобрать наиболее подходящую иерархию бывает затруднительно. Проблему можно решить так: достаточно поставить несколько тегов для всех изображений, а сами они могут храниться в плоской системе файлов. Благодаря этому свойству тегирование используется для рубрикации контента не только только онлайн, но и в оффлайн приложениях, например, просмотрщиках фотографий.

Можно выделить две разновидности популярных хэштегов в социальных сетях. Первые, используемые недолго и посвященные каким-то социальным явлениям или событиям, например: *#elections2018*, *#metoo*. И вторые, широкораспространенные, но не связанные с новостной повесткой, например: *#sport*, *#cafe*; они и будут нас интересовать. Данная работа посвящена разработке интеллектуальной системы, подсказывающей пользователю релевантные хештеги к загружаемым фотографиям. Кроме того, с помощью такой системы можно решать и “обратную” задачу — определять, уместно ли поставлены те или иные теги к заданным изображениям. Способность системы давать ответ на такой вопрос можно использовать для выявления злоупотреблений со стороны пользователей. Например, зачастую в рекламных целях продвигаемые публикации снабжают множеством популярных тегов, не имеющих никакого отношения к публикуемой информации.

1 Постановка задачи

Целью настоящей работы является построение интеллектуальной системы для структурирования изображений в социальных сетях. А именно, предлагается автоматизировать процесс добавления пользовательских хештегов к загружаемым изображениям. С точки зрения машинного обучения решается задача классификации изображений.

В настоящее время наилучшие результаты в задаче классификации изображений удаётся получить, используя глубокие сверточные нейронные сети. В качестве примера можно привести один из самых известных и больших конкурсов по классификации изображений датасета *ImageNet*[1], который проводится ежегодно с 2010 года. В настоящее время решения всех призеров так или иначе базируются на сверточных нейронных сетях. [TODO links](#)

Идея применить машинное обучение к тегированию изображений в социальных сетях не нова, можно привести в пример исследования [s1](#), [s2](#) и [s3](#) [TODO links](#).

В данной работе предлагается способ улучшить точность предсказания популярных тегов, сузив информационный домен, к которому эти теги относятся. Большая часть из наиболее употребимых тегов в социальных сетях описывает эмоции, чувства или другие абстрактные понятия, не имеющие прямого выражения в объектах реального мира. Например, в 2018 году одними из самых популярных хештегов в сети *Instagram* стали *#love*, *#happy*, *#beautiful*. Понятно, что модели компьютерного зрения наоборот будут точнее работать для тегов, связанных с наличием в кадре тех или иных сущностей, например *#beach*, *#sky* или *#architecture*. Для таких случаев и будет строиться модель, описанная в настоящей работе.

В качестве датасета, который может быть использован для наших целей и разметка которого напрямую связана с объектами, находящимися в кадре, был выбран *Scene Understanding Dataset (SUN)*. Это набор изображений для каждого из которых выбрана одна из четырехсот локаций (сцен), вот несколько примеров:

- *baseball field*
- *basketball court*
- *ice shelf*
- *forest*
- *wind farm*

Большинство названий локаций сами по себе не являются популярными тегами из социальных сетей. Поэтому необходимо сопоставить их широко распространенным хештегам (если это возможно):

- *baseball field, basketball court* → *#sport*

- *ice shelf, forest* → *#nature*
- *wind farm* → ?

Выполнив сопоставление, можно присутствовать к написанию и обучению сверточной нейронной сети. В итоге будет получена модель, которая по окружению, обнаруженному на пользовательском фото, сможет подсказывать подходящий хэштег.

Ясно, что полученная модель будет корректно работать лишь для ограниченного (пусть и большого) домена фотографий. Следовательно, необходимо обучить её распознавать не входящие в этот домен изображения и не пытаться определить их категорию. Кроме того, в случае низкой уверенности в правильности предсказания так же лучше ничего не делать. По мнению автора, гораздо предпочтительнее не предложить пользователю подходящий хэштег, чем многократно предлагать нерелевантные варианты.

Наконец, точность предсказаний обученной модели будет проверена вручную группой пользователей на выборке реальных фотографий из социальных сетей.

Научная новизна работы определяется:

- Адаптацией датасета *SUN* для решения задачи о структурировании изображений в социальной сети;
- Исследованием новых схем обучения нейронных сетей, предложенных автором.

2 Используемые инструменты

Программные средства

В качестве языка программирования использовался *Python 3.6.7* (сборка *Anaconda*), в качестве среды разработки — *PyCharm Professional 2018.1*, операционная система *Ubuntu 16.04.4 LT*. Использованы следующие сторонние библиотеки для *python*:

- *pytorch, torchvision* — построение и обучение нейронных сетей
- *tensorboardX* — визуальное логирование процесса обучения
- *PIL, opencv, scikit-image, scipy* — обработка изображений
- *matplotlib* — отрисовка графиков
- *numpy* — матричные вычисления
- *pandas* — работа с таблицами
- *nltk* — работа с текстом, в том числе с базой *WordNet*
- *scikit-learn* — библиотека машинного обучения общего плана
- *pip* — пакетный менеджер

Вычислительные мощности

Обучение моделей производилось на удаленном сервере со следующей конфигурацией:

- видеокарта *GEFORCE GTX 1080 Ti*, (11 ГБ видеопамяти)
- процессор *AMD Ryzen Threadripper 1920X 12-Core*
- оперативная память объемом 62 ГБ

3 Сверточные нейронные сети

3.1 Введение в теорию сверточных нейросетей

3.2 Обзор используемых архитектур

4 Подготовка данных

4.1 База WordNet

4.2 Датасет SUN

4.3 Адаптация датасета SUN

5 Численные эксперименты

5.1 Архитектура программы

5.2 Особенности реализации

5.3 Полученные результаты

5.4 Обсуждение результатов

6 Валидация результатов

Выводы

Список литературы

- [1] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. [2015] *ImageNet Large Scale Visual Recognition Challenge*. IJCV.