

Automated Video Captioning

Ткаченко Арэс

Хачатрян Саргис



Цели задачи

- В задаче Video Captioning (создание текстового описания к видео) модели необходимо проанализировать короткий видеофрагмент и сгенерировать наиболее подходящее текстовое описание на английском языке, которое характеризует события и/или действия, происходящие на видео.

	index	file_name	caption
0	0	0.mp4	A man is working out on a seated chest press m...
1	1	1.mp4	Preparing a bowl with yogurt and assorted fres...
2	2	2.mp4	A man with a muscular build is seen from behin...
3	3	3.mp4	Man exercising by jogging on a pedestrian brid...
4	4	4.mp4	Wristwatch hands moving forward close-up views.

Проблемы датасета

- **Данные - видео:** т.е. их нужно как-то разбивать на кадры
- **Много длинных описаний:**
- **Малое-кол-во данных:**

Модели:

- Так как текст и видео это последовательные наборы данных, мы решили использовать рекуррентные нейронные сети и трансформер
- Мы остановились на 2 моделях:
 - 1) Рекуррентная модель с resnet и lstm
 - 2) Transformer

Подходы:

Каждую из нейросетей мы обучили на тренировочных данных, предварительно их обработав, после чего проверили на тестовых данных и уже получили некоторый результат

Результаты и метрики

Результат мы получали за счет метрики BLEU, эта метрика смотрит схожесть полученного текста и того что мы предсказали, учитывая при этом вариативность

Automated Video Captioning

Submit Prediction

OverviewDataCodeModelsDiscussionLeaderboardRulesTeamSubmissions

AllSuccessfulSelectedErrorsRecent

Submission and Description

Public Score ⓘSelect

✓bumblebee.csv

Complete · 22s ago

0.00723

✓optimus.csv

Complete · 1m ago

0.00906

✓submission.csv

Complete · 12h ago

0.00292

Данная метрика показала нам 1%, вероятнее всего это связано с тем что у нас был не очень большой набор тренировочных данных, а также достаточно тяжело при сравнении текстов добиться большого показателя

Сравнение моделей

Метрика: Lstm и resnet: 0,007

Transformer: 0,009

Модель Transformer получила большую метрику так как она более подстроен для подобных задач, но требует большего объема памяти, а также времени



Идеи для улучшения

- **1. Балансировка датасета**

Аугментация существующих видео: зеркальное отражение, изменение скорости


- **Низкие значения BLEU:**

Полученные значения BLEU — 0.009 для модели OptimusPrime и 0.007 для Vumblebee — указывают на то, что качество сгенерированных описаний видео на текущем этапе остаётся очень низким. Возможно причина кроется в малом кол-ве данных и небольшом числе эпох обучения. Т.е. стоит обучать модели дольше.

- **2. Улучшение обработки видео**

Использование большего числа кадров (с 16 до 32–64) для лучшего анализа движений

Выбор ключевых кадров вместо равномерной выборки



Презентация закончилась

окак