УДК 004.852

ОПРЕДЕЛЕНИЕ НАИБОЛЕЕ ПОДХОДЯЩЕГО ФОНА ПО ОПИСАНИЮ ИЗОБРАЖЕНИЯ

Чижиков Д.А. (Национальный исследовательский университет ИТМО)
Научный руководитель – к.ф.-м.н., доцент ФИТиП Фильченков А.А.
(Национальный исследовательский университет ИТМО)

В работе описано решение задачи – по данному текстовому описанию определить место, в котором происходит описанное в нем действие. В работе исследуются существующие решения в этой области и производится собственное обучение модели-трансформера.

Введение.

Часто текст для большего понимания сопровождается иллюстрацией, при этом задача автоматической генерации изображения по тексту является сложной задачей компьютерного зрения. В последнее время появилось много видов порождающих состязательных сетей (Generative Adversarial Networks, GANs), решающих её, но они генерируют нереалистичные изображения низкого качества.

Данную задачу можно разбить на несколько подзадач, в числе которых будет нахождение ключевых объектов текста и их описаний, но при этом в описании часто не содержится информация о том, где происходит действие. Конкретного решения у данной задачи нет, поэтому необходимо реализовать собственное, а для этого нужно:

- Составить синтетический набор данных.
- Обучить модель-трансформер на текстовых описаниях из составленного набора данных.
- Сравнить результаты с существующими решениями аналогичных задач.

Основная часть.

В первую очередь для тестирования и сравнения результатов был составлен набор данных из 300 коротких текстов 20-ти классов, которые разбиты на два поднабора по 250 и 50 текстов 20 и 12 классов соответственно для тестирования выделения описания из текста, если оно в нем есть, и предсказания, если его в нем нет.

На основе набора данных MS COCO собран синтетических набор данных из 30000 объектов 3000 классов путем обработки описаний изображений и применения алгоритма Image Captioning.

Перед обучением модели-трансформера было принято решение проверить существующие решения аналогичных задач, а именно:

- Алгоритм EmbedRank – алгоритм для извлечения ключевых слов, который применяет эмбеддинги для получения информации о различиях между текстом и кандидатами на ключевые слова. В качестве эмбеддингов рассмотрены:
  - FastText, а именно одна из его версий – Sent2Vec.
  - Различные версии эмбеддинга BERT.
- Распознавание именованных сущностей. В качестве фона может выступать именованный объект, выделение именованных объектов относится к задаче распознавания именованных сущностей, решения которой предоставляют такие наборы библиотек для обработки естественного языка как spaCy[1], nltk[2] и natasha[3].

1 https://spacy.io/
2 https://www.nltk.org/
3 https://natasha.github.io/ner/

- Частеречная разметка, которую также могут осуществлять spaCy и nltk.
- GPT-2 – генеративная нейронная сеть для предсказания последовательности на основе данного ей текста.

В дальнейшем собранный синтетический набор данных будет использован для обучения модели-трансформера, которая будет протестирована на наборе данных для тестирования, результаты тестирования будут сравнены со всеми описанными решениями.

Выводы.

После сравнения результатов лучший алгоритм будет реализован в сервисе на Python, который будет способен выделять (если оно есть) или предсказывать (если его нет) описание места действия в данном коротком тексте на английскому или русском языке.

Чижиков Д.А. (автор) Подпись

Фильченков А.А. (научный руководитель) Подпись