

Рецензия на работу “Анализ аргументации в отзывах на кинофильмы”

В проекте “Анализ аргументации в отзывах на кинофильмы” было проведено исследование способности языковых моделей решать задачу классификации позиций и доводов. Авторы проекта использовали идею соревнования RuArg-2022 для создания собственного датасета из отзывов пользователей сайта Кинопоиск. Была осуществлена ручная разметка части данных из датасета blinoff/kinopoisk: размечены доводы, присутствующие в отзывах, и дополнена разметка позиций. При составлении нового датасета авторы постарались соблюсти принципы разметки, использованные в RuArg-2022, а также сохранить соответствующее соотношение классов. После этого на полученном датасете были обучены две энкодерные модели: DeepPavlov/rubert-base-cased и ai-forever/ru-Roberta-large.

Итоговое качество обученных моделей оказалось достаточно высоким: удалось добиться F-меры 72.37% для классификации утверждений и 77.8% для классификации доводов, что выше максимальных результатов, полученных в соревновании RuArg-2022. Гипотезы, выдвинутые авторами проекта, частично подтвердились: качество моделей при анализе новых фильмов оказалось выше, чем при анализе старых, модель классифицировала утверждения о фильмах из списка лучших правильнее, чем о фильмах из списка худших.

Помимо представленных результатов обучения, было бы интересно увидеть их обсуждение и анализ, в том числе информацию о том, подтвердилась ли гипотеза о большей частоте предсказаний конгруэнтных позиций и доводов на месте неконгруэнтных по сравнению с предсказаниями неконгруэнтных позиций и доводов на месте конгруэнтных. Возможно, существуют факторы, объясняющие разные результаты двух моделей при обучении на полученном датасете.

В заключение хочется отметить большой объем проделанной работы: для составления датасета авторами проекта были проанализированы и размечены 3280 отзывов. Большим плюсом является также подробный анализ структуры как датасета RuArg-2022, так и датасета отзывов с Кинопоиска.