

Анализ аргументации в отзывах на кинофильмы

Виктория Борисова
Инна Рабинович
Анна Сербина

ОТиПЛ МГУ имени М. В. Ломоносова

Определение

Анализ аргументации (*argumentation mining*) – это область компьютерной лингвистики, в которой исследуются методы извлечения из текстов и классификации аргументов и связей между ними, а также построения аргументационной структуры

Аргумент = утверждение (claim) + довод (premise)

- **Утверждение** выражает позицию («за» или «против»)
- **Довод** – доказательства / мотивация для подтверждения позиции

Пример: RuArg-2022 – задача анализа аргументации относительно COVID-19

Структура

Поля

- `text_id` – номер текста
- `text` – текст
- `masks_stance` – позиция по маскам
- `masks_argument` – доводы по маскам
- `quarantine_stance` – позиция по карантину
- `quarantine_argument` – доводы по карантину
- `vaccines_stance` – позиция по вакцинам
- `vaccines_argument` – доводы по вакцинам

В 3 текстах тренировочного датасета есть мнение по трем темам, в 88 – по маскам и карантину, в 55 – по маскам и вакцинам, в 31 – по карантину и вакцинам. В остальных 6540 текстах есть мнение только по одной теме

Состав

train	6717 (70%)
validation	1431 (15%)
test	1402 (15%)
Σ	9550 (100%)

Соотношение классов (маски)

позиция / довод	за (2)	нет арг. (1)	против (0)	Σ
за (2)	275 (9%)	428 (13%)	1 (0%)	704 (22%)
другое (1)	61 (2%)	1713 (55%)	58 (2%)	1832 (59%)
против (0)	3 (0%)	310 (10%)	281 (9%)	594 (19%)
Σ	339 (11%)	2451 (78%)	340 (11%)	3130 (100%)

Соотношение классов (карантин)

позиция / довод	за (2)	нет арг. (1)	против (0)	Σ
за (2)	180 (8%)	405 (20%)	2 (0%)	587 (28%)
другое (1)	37 (2%)	1247 (59%)	57 (3%)	1341 (64%)
против (0)	0 (0%)	104 (5%)	68 (3%)	172 (8%)
Σ	217 (10%)	1756 (84%)	127 (6%)	2100 (100%)

Соотношение классов (вакцины)

позиция / довод	за (2)	нет арг. (1)	против (0)	Σ
за (2)	125 (8%)	249 (15%)	0 (0%)	374 (23%)
другое (1)	21 (1%)	791 (48%)	54 (3%)	866 (52%)
против (0)	3 (0%)	198 (12%)	217 (13%)	418 (25%)
Σ	149 (9%)	1238 (75%)	271 (16%)	1658 (100%)

Исследования с IMDb

- Работа с датасетом IMDb (Internet Movie Database) с помощью разных методов машинного обучения [[Handayani 2023](#)], [[Talibzade 2023](#)]

Наиболее популярная задача – анализ тональности. Используются tf-idf векторизация, логистическая регрессия и метод опорных векторов

- Исследование, посвященное связи между упоминанием фильма в соцсетях и его рейтингом на IMDb [[Oghina et al. 2012](#)]

В этой работе в частности использовался подход, принимающий во внимание лексику в записях соцсетей

Датасет blinoff/kinopoisk

Для Кинопоиска был создан датасет blinoff/kinopoisk:

<https://huggingface.co/datasets/blinoff/kinopoisk>

- Использовался на Российском семинаре по Оценке Методов Информационного Поиска для дополнительного обучения модели и проверки гипотез [Blinov et al. 2013]
- Был применен лексический метод и методы машинного обучения
- У каждого подхода свои плюсы и методы, поэтому наилучших результатов можно достичь их совокупностью

Структура

Поля

- part: top250 или bottom100
- movie_name – название фильма
- review_id – id рецензии
- author – автор рецензии
- date – дата написания рецензии
- title – название рецензии
- grade3 – отрицательная / положительная / нейтральная
- grade10 – оценка из 10
- content – текст рецензии

Содержит 36591 рецензию на фильм с сайта «Кинопоиск» (июль 2004 – ноябрь 2012), разделения на на train, validation и test нет. Всего описано 350 фильмов: топ-250 лучших и топ-100 худших по оценкам сайта



Кинопоиск

24 сентября 2011

Плакали наши денежки (с)

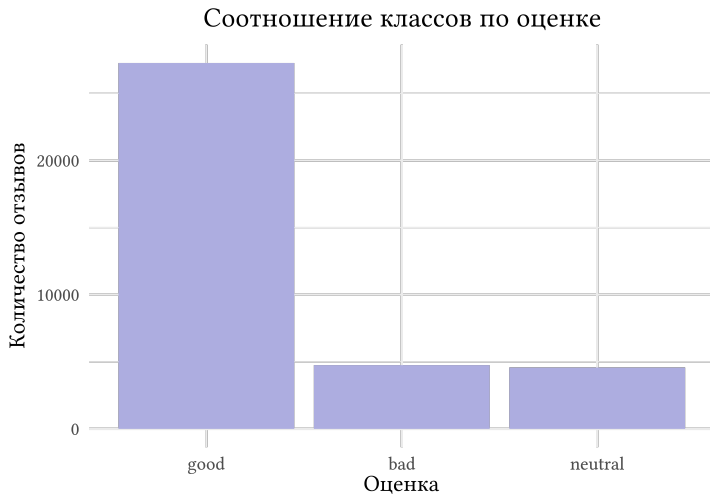
'Блеф' - одна из моих самых любимых комедий.

Этот фильм я наверно смотрел раз сто, нет я конечно блефую, я видел его куда больше. Не могу не выразить своё восхищение главными действующими лицами этого фильма. Начну с Адриано Челентано для которого как я считаю это лучшая роль в кино. Великолепный актёр, неплохой певец, странно что на его родине в Италии его песни мало кто слушает. Ну я думаю что и итальянцы и французы привыкли к тому, что у нас до сих их актёры популярней чем даже на своей родине. Да, такой вот парадокс.

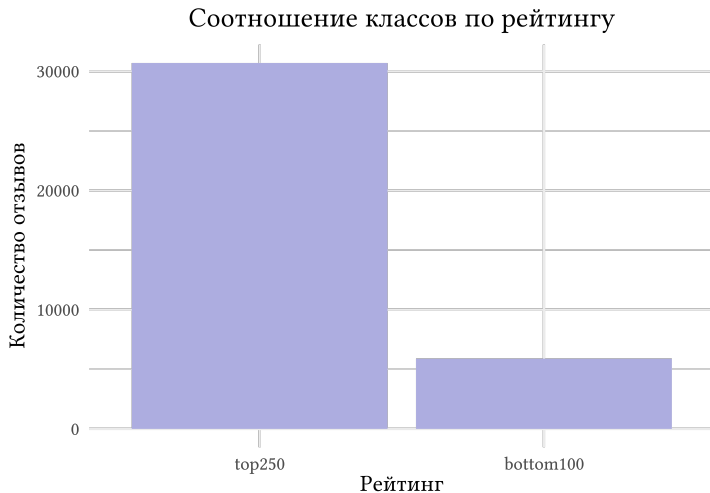
Челентано конечно профессионал своего дела, комик с серьёзным выражением лица. Он смешон ещё и потому, что одновременно так серьёзен. Адриано браво!...

part	movie_name	review_id	author	date	title	grade3	grade10	content
string · classes	string · classes	string · lengths	string · lengths	timestamp[s]	string · lengths	string · classes	string · classes	string · lengths
2 values	350 values	1	1	34 s	1	3 values	69 values	18.5k
top250	Блеф (1976)	17144	Come Back	2011-09-24T00:00:00	Плакали наши денежки ☹	Good	10	"Блеф" — одна из моих самых любимых комедий. Этот фильм...
top250	Блеф (1976)	17139	Stasiki	2008-03-04T00:00:00	null	Good	8	Адриано Челентано продолжает радовать нас своими работами...
top250	Блеф (1976)	17137	Flashman	2007-03-04T00:00:00	null	Good	10	Несомненно, это один из великих фильмов 80-х годов.
top250	Блеф (1976)	17135	Sergio Tishin	2009-08-17T00:00:00	* Черное, красное, ерунда это все...	Good	8	Эта фраза на мой взгляд отражает сюжет несомненно...
top250	Блеф (1976)	17151	Фелъгья	2009-08-20T00:00:00	«Он хотел убить?» Да! Блеф, блеф...	Neutral	7	- как пела Зенфира, скорее всего, по совершенно другому...
top250	Блеф (1976)	17142	Marvel	2005-04-06T00:00:00	null	Good	8	Беспорный классический шедевр, который и через 20...
top250	Блеф (1976)	17147	NaZ LemberG	2012-03-31T00:00:00	Меня всегда убеждали убедительные доводы!	Good	9	- Ну, так что ты скажешь по поводу нашей совместной...
top250	Блеф (1976)	17150	Horseehell	2011-11-15T00:00:00	null	Neutral	7	Финансовая акула Бель Дек имеет давние счёты...
top250	Блеф (1976)	17148	MC LOUD	2012-03-19T00:00:00	null	Good	8	Весьма забавное и легкое кино. Челентано как всегда в своем...
top250	Блеф (1976)	17134	Дегустатор	2010-02-18T00:00:00	null	Good	10	Наверное, всё-таки самый лучший фильм с участием...
top250	Блеф (1976)	17146	marina357	2011-01-14T00:00:00	Блеф по-итальянски	Good	9	«Блеф» — комедия Серджио Корбуччи 1976 года. Один...
top250	Блеф (1976)	17149	Годлев Дмитрий	2008-11-10T00:00:00	«жук с права!приятно узнать... А это у меня...	Good	8	Действие фильма разворачивается примерно...

Состав (оценки)



Состав (рейтинг)



Гипотезы

- Позиция и аргументация будут лучше предсказываться для современных фильмов (XXI века)
- Позиция и аргументация будут лучше предсказываться для фильмов из вершины рейтинга
- Конгруэнтные позиции и аргументации (например, *Good* и *Good*) могут предсказываться лучше, чем неконгруэнтные

Задачи и метод

Задачи

- Определить подходящий размер датасета и соотношение классов в нем
- Осуществить ручную разметку доводов
- Провести дообучение модели на обучающей выборке
- Оценить результаты работы модели на валидационной и тестовой выборках
- Провести анализ результатов в соответствии с поставленными гипотезами

Метод

- Дообучение энкодерных моделей

Какой датасет нам нужен?

- Необходимый объем: train – 2296 (70%), validation – 492 (15%), test – 492 (15%)

В тренировочном датасете RuArg-2022 в среднем 2296 текстов на тему

- Сбалансированное соотношение позиций

- N самых коротких рецензий

Это сблизит нас с датасетом RuArg-2022 (например, в blinoff/kinopoisk 3242 рецензии, чья длина < 94 слов)

- Инструкция для разметчиков доводов с RuArg-2022 и терпение :)

Какая модель нам нужна?

- DeepPavlov/rubert-base-cased

С ее помощью было создано базовое решение организаторов RuArg-2022

- ai-forever/ru-Roberta-large

Учитываем опыт исследований, связанных с IMDb

Разметка данных

При разметке использовались принципы разметки RuArg-2022

- В каждом отзыве подсчитывалось количество аргументов «за» и количество аргументов «против»
- На основании полученных чисел выставлялся лейбл:
 - если в обеих колонках «0» – Neutral
 - если в одной колонке «0» – лейбл, соответствующий колонке с ненулевым значением
 - если в обеих колонках ненулевые значения и разность двух чисел меньше 3 – Neutral
 - если в обеих колонках ненулевые значения и разность чисел не меньше 3 – лейбл, соответствующий колонке с наибольшим значением
- Так как для старых отзывов категория Neutral выставлена по умолчанию, часть из них была переразмечена

Наш датасет можно увидеть по ссылке:

https://huggingface.co/datasets/otipl2125/film_review_argumentation

Соотношение классов полученного датасета

позиция / довод	good	neutral	bad	Σ
good	1130 (34%)	247 (8%)	32 (1%)	1409 (43%)
neutral	178 (5%)	514 (16%)	59 (2%)	751 (23%)
bad	47 (1%)	398 (12%)	675 (21%)	1120 (34%)
Σ	1355 (40%)	1159 (36%)	766 (24%)	3280 (100%)

Почему аргументация не всегда соответствует категории отзыва?

- Категория отзыва выставлена на основе субъективной оценки, а в отзыве приведены контраргументы. Пример: «Я в восторге от фильма, но CGI нереалистичный. Все равно 9.5/10»
- Какой-то фактор(-ы) оказывается для пользователя важнее всех остальных. Пример: «Прекрасная постановка сцен, красивая картинка, продуманный нешаблонный сюжет, невероятно мелодичная музыка. Единственный недостаток – главная героиня с ее наивным характером и глупыми выходками. Не проникся.»

Почему аргументация не всегда соответствует категории отзыва?

- Пользователь приводит аргументы других людей/контраргументы
- В отзыве нет аргументации, он сводится к пересказу сюжета или личным историям, связанным с просмотром фильма
- Фильм является общепризнанным, но не понравился пользователю, или же фильм напротив имеет низкие оценки, но пользователь получил удовольствие. В обоих случаях может получиться нейтральная рецензия

Дообучение моделей

- Перевод датасета в необходимый формат
 - Добавление столбцов
 - Век
 - Принадлежность к лучшей / худшей части списка
 - Конгруэнтность
 - Tokenизация отзывов
 - Разбиение по батчам
- Добавление двух слоев для улучшения работы моделей
 - Для классификации меток отзывов
 - Для меток аргументации
- Обучение обеих моделей ([DeepPavlov/rubert-base-cased](#) и [ai-forever/ru-Roberta-large](#)), 2 эпохи

Результаты (F1-мера)

	DeepPavlov/rubert-base-cased	ai-forever/ru-Roberta-large
Stance	72.37%	70.79%
Argument	72.68%	77.8%

Результаты по веку (F1-мера)

	DeepPavlov/rubert-base-cased	ai-forever/ru-Roberta-large
Stance	70.17% (XX век)	63.62% (XX век)
	72.16% (XXI век)	72.25% (XXI век)
Argument	60.27% (XX век)	66.22% (XX век)
	74.18% (XXI век)	79.83% (XXI век)

Примечание по соотношению данных: отзывов по фильмам XX века – 1008 (160 в тестовой выборке), по фильмам XXI века – 2272 (332 в тестовой выборке)

Результаты по рейтингу (F1-мера)

	DeepPavlov/rubert-base-cased	ai-forever/ru-Roberta-large
Stance	71.63% (хорошие)	68.23% (хорошие)
	62% (плохие)	61.08% (плохие)
Argument	66.18% (хорошие)	74.84% (хорошие)
	68.56% (плохие)	76.76% (плохие)

Примечание по соотношению данных: отзывов по 250 лучшим фильмам – 2226 (352 в тестовой выборке), по 100 худшим фильмам – 1054 (140 в тестовой выборке)

Результаты по конгруэнтности (F1-мера)

	DeepPavlov/rubert-base-cased	ai-forever/ru-Roberta-large
Stance	79.64% (конгр.)	80.57% (конгр.)
	50.39% (неконгр.)	49.37% (неконгр.)
Argument	79.78% (конгр.)	83.36% (конгр.)
	47.98% (неконгр.)	60.09% (неконгр.)

Примечание по соотношению данных: конгруэнтных отзывов – 2319 (339 в тестовой выборке), неконгруэнтных – 961 (153 в тестовой выборке)

Выводы

- По гипотезе 1: позиция и доводы лучше предсказываются для фильмов XXI века
- По гипотезе 2: позиция лучше предсказывается для «хороших» фильмов, с доводами все наоборот
- По гипотезе 3: позиция и доводы лучше предсказываются, если они совпадают

Литература

Blinov P. D., Klekovkina M. V., Kotelnikov E. V., Pestov O. A. (2013)
Research of lexical approach and machine learning methods for sentiment analysis
Computational Linguistics and Intellectual Technologies, 2(12). P. 48-58.

Handayani T. (2023)
Unmasking the Hidden Emotions: Sentiment Analysis on IMDb Movie Reviews
URL: <https://medium.com/@mbaktrihandayani/unmasking-the-hidden-emotions-sentiment-analysis-on-imdb-movie-reviews-9573779c304c>

Oghina A., Breuss M., Tsagkias M., Rijke M. (2012)
Predicting IMDB Movie Ratings Using Social Media
Advances in Information Retrieval. ECIR 2012. P. 503-507.

Talibzade R. (2023)
Sentiment Analysis of IMDb Movie Reviews Using Traditional Machine Learning Techniques and Transformers.
Term paper, ADA University, Baku.

Спасибо за внимание!