

Структурированный анализ тональности: данные и методы

Лукашевич Н.В.

НИВЦ МГУ имени М.В. Ломоносова

Анализ тональности представляет собой одно из наиболее активно развивающихся направлений в автоматической обработке текстов. В начале исследований в этой области ставились задачи определения общей тональности отзывов пользователей или общей тональности сообщения в социальной сети. В настоящее время имеется большое разнообразие постановок задач, включая определение тональности по отношению к конкретной сущности, аспекту (части или характеристике) этой сущности, близкая задача определения позиции (stance) по некоторой дискуссионной теме. Современные методы позволяют приблизиться к более полному извлечению мнения (Liu, 2012), включая источник мнения, объект мнения, тональность и языковое выражение, которым выражена данная тональность. Именно такая задача извлечения мнений ставилась для открытого тестирования в рамках семинара SemEval-2022 (Task 10, Structured sentiment analysis) (Barnes et al., 2022).

Для русского языка создается корпус для структурированного анализа тональности RuSentNE, который состоит из новостных текстов. Источником мнения может быть автор текста, автор цитаты или другая упоминаемая сущность. Мнение может быть имплицитным, т.е. выражаться через действия сущности, например, в предложении “X уволил Y” предполагается, что имеется имплицитное негативное мнение X по отношению Y.

Анализ тональности новостных текстов – важное направление в области анализа мнений, поскольку обнаружение, отслеживание трендов тональности в новостном потоке важно для построения разного рода аналитических систем, мониторинга имиджа в СМИ конкретных людей или компаний. При этом данная постановка задачи недостаточно исследована, что связано с высокой сложностью извлечения тональности. Новостные тексты, с одной стороны, содержат высокую долю нейтрально упомянутых сущностей. С другой стороны, новостной текст может содержать противоположные мнения по отношению к одной и той же сущности (Loukachevitch, Rusnachenko, 2018).

Первая версия создаваемого корпуса стала основой для открытого тестирования RuSentNE-2023, в котором требовалось определение тональности по отношению к заданной именованной сущности, т.е. нужно было предсказать, позитивно или негативно содержание заданного предложения по отношению к заданной сущности (Golubev et al., 2023). Тестирование было успешно проведено, на площадке тестирования <https://codalab.lisn.upsaclay.fr/competitions/9538> до сих пор продолжают появляться новые результаты. Наилучшими подходами к анализу тональности в 2023 году были ансамбли методов на основе модели BERT (Devlin et al., 2019). В настоящее время лучший результат в данной постановке задачи достигнут на основе обученной на модели FLAN-T5 (Rusnachenko et al., 2024).

Рисунок 1 показывает пример разметки в корпусе RuSentNE. Данная разметка будет выгружена в соответствующие кортежи мнений для проведения экспериментов. В частности, на основе этого примера должен быть порожден следующий кортеж мнения: (Маттео Ренци, Фо, pos, крупнейшую фигуру культурной жизни), т.е. Маттео Ренци позитивно относится к Фо, и это позитивное мнение отражено в словах «крупнейшую фигуру культурной жизни».

В докладе будет представлен корпус RuSentNE, а также результаты работы моделей по извлечению мнений на этом корпусе.

Список литературы

Barnes J. et al. SemEval-2022 task 10: Structured sentiment analysis. Proceedings of the 16th International Workshop on Semantic Evaluation. 2022.

Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019. С. 4171-4186.

Golubev A. et al. RuSentNE-2023: Evaluating Entity-Oriented Sentiment Analysis on Russian News Texts. Proceedings of the International Conference "Dialogue. 2023. Т. 2023.

Liu B. Sentiment analysis and opinion mining. Springer Nature, 2012.

Loukachevitch N., Rusnachenko N. Extracting sentiment attitudes from analytical texts //Komp'juternaja Lingvistika i Intellektual'nye Tehnologii. 2018. С. 448-458.

Rusnachenko N., Golubev A., Loukachevitch N. Large Language Models in Targeted Sentiment Analysis //arXiv preprint arXiv:2404.12342. 2024.



Рисунок 1. Пример разметки в корпусе RuSentNe