

Применение трансферного обучения для анализа тональности текста

Матвеев Данила Денисович

Студент (бакалавр)

Институт медицины, экологии и физической культуры, Россия

E-mail: danilaconcyrc@gmail.com

В современную цифровую эпоху, когда ежесекундно генерируются огромные объемы текстовых данных на различных онлайн-платформах, понимание смысла этих текстов приобретает решающее значение[1, 3]. Анализ тональности - область на стыке лингвистики

и машинного обучения - призван автоматизировать процесс определения выраженных в тексте настроений, начиная от положительных и отрицательных мнений и заканчивая нюансами эмоций[2].

Данная статья посвящена анализу тональности текста и сосредоточена на ключевой роли, которую методы машинного обучения играют в обработке естественного языка. Анализ тональности имеет огромное значение в различных областях, включая маркетинг, анализ отзывов клиентов, мониторинг социальных сетей, анализ политических лозунгов и многое другое. Используя алгоритмы машинного обучения, анализ настроений позволяет организациям извлекать ценные сведения, принимать решения на основе данных и эффективно адаптировать свои ответы к потребностям и предпочтениям целевой аудитории.

Для анализа предлагается обучение и использование нейросети T5(Transformers)[4,5] на объединенном наборе данных, собранным из материалов с сайтов Reuters, X(Twitter), Reddit,

Amazon, IMDb общим объемом в ~100000 уникальных записей. Каждый материал(статья/комментарий) был вручную размечен одной из трех меток (“позитивное”, “нейтральное” или “негативное” высказывание).

Далее нейросеть ruT5-base[6] была обучена на кластере NVIDIA A100, после чего был развернут веб-интерфейс на базе Streamlit для проведения тестов.

В заключение следует отметить, что применение моделей на основе Transformers в анализе тональности представляет собой значительное достижение в области обработки естественного языка. Благодаря анализу огромных объемов текстовых данных, модель ruT5

продемонстрировала замечательные возможности в понимании и интерпретации человеческих чувств, выраженных в тексте.

Становится очевидным, что подходы на основе Transformers и GPT обладают беспрецедентной гибкостью, масштабируемостью и адаптивностью. Эти модели превосходно улавливают тонкие нюансы, контекстуальные зависимости и лингвистические хитросплетения, присущие человеческому общению, повышая тем самым точность и глубину анализа настроений.

Источники и литература

- 1) Акжолов Р. К., Верига А. В. Анализ тональности текста // Вестник науки. – 2019. – Т. 4. – №. 9 (18). – С. 29-31.
- 2) Котельников Е. В., Клековкина М. В. Автоматический анализ тональности текстов на основе методов машинного обучения // Компьютерная лингвистика и интеллектуальные технологии. – 2012. – Т. 2. – №. 11. – С. 27.

- 3) Маньков А. С., Детков А. А. Анализ тональности текста. Обзорная статья // Весенние дни науки: сборник докладов Международной конференции студентов и молодых ученых. — Екатеринбург, 2023. — Издательство Издательский Дом «Ажур», 2023. — С. 66-68.
- 4) Raffel C. et al. Exploring the limits of transfer learning with a unified text-to-text transformer // Journal of machine learning research. — 2020. — Т. 21. — №. 140. — С. 1-67.
- 5) Waswani A. et al. Attention is all you need // NIPS. — 2017.
- 6) Zmitrovich D. et al. A family of pretrained transformer language models for Russian // arXiv preprint arXiv:2309.10931. — 2023.

Иллюстрации

Примеры анализа

Платформа	Текст	Оценка тональности
Reuters	"US to impose new visa curbs on Hong Kong officials over rights crackdown"	Negative
X(Twitter)	"This next Superbowl is on my (21st) birthday. Random fact of the day."	Neutral
X(Twitter)	"Woot! So excited that I get to watch tonight's game. Go Colts!"	Positive
IMDB	"The movie is very realistic. Absolutely, it does not belong to the Hollywood Cinema genre where every line must be pronounced in a perfect manner and where every move is precise. The actors playing the roles of the lovers do a GREAT job representing the characters' feelings and thoughts - their everyday life adventures. Overall, the movie climaxes the viewer to a depressed state. This is where the realism of the whole story is apparent. Not everything happens the way we think it should happen. I can say though that the movie does not end on a bad note. We watch, we learn, we experience ourselves. That is probably the moral of this story."	Positive

Рис. : Примеры анализа