

ВЫЯВЛЕНИЕ И АНАЛИЗ ЭМОЦИОНАЛЬНОГО КОНТЕКСТА ЦИФРОВЫХ ТЕКСТОВ

В условиях постоянного роста объемов цифровых данных, в том числе текстовой информации, определение эмоционального фона текста становится одной из приоритетных задач цифровой экономики. Выявление и анализ эмоционального контекста в текстах, или анализ тональности текста (сентимент-анализ), является новым активно развивающимся направлением в области автоматической обработки естественно-языковых текстов, которое предполагает определение эмоциональной окраски сообщений, выявления эмоционально окрашенной лексики и эмоциональной оценки объектов. Это требует не только выявления базовых эмоций, но также учета контекста и языковых особенностей текстовых конструкций.

В современном мире анализ эмоциональной окраски текста приобретает все большее значение в различных прикладных областях. Одним из важнейших применений является мониторинг психического здоровья населения через анализ публичных сообщений в социальных сетях, чатах, форумах и т.п. На основе лингвистических маркеров в текстах пользователей системы автоматического мониторинга направлены на выявление признаков депрессии, агрессивного поведения, суицидальных наклонностей или иных нетипичных форм поведения. Это позволит своевременно идентифицировать людей, которые нуждаются в психологической помощи, и предотвратить возможные трагические события.

В сфере маркетинга и бизнес-аналитики анализ эмоциональной окраски помогает оценить отношение потребителей к продуктам и услугам. Автоматическая обработка отзывов, комментариев и упоминаний в социальных медиа позволяет отслеживать репутацию бренда, выявлять проблемные аспекты продуктов и оперативно реагировать на негативные тенденции.

В области безопасности анализ эмоциональной окраски текста поможет выявить потенциальные угрозы, включая экстремистские настроения или подготовку противоправных действий. Решение данной задачи требует мониторинга больших объемов текстовой информации и возможностей для выявления паттернов агрессивного или

деструктивного поведения. Это особенно важно для защиты уязвимых групп населения (общение подростков в социальных сетях).

Кроме того, изучение способов выражения эмоций в языке имеет большое значение для понимания межкультурной коммуникации, так как разные культуры могут иметь различные традиции и нормы эмоционального выражения. Это находит отражение в лексическом составе языков: некоторые эмоциональные состояния могут иметь специальные обозначения в одном языке и отсутствовать в другом.

В зависимости от контекста и природы эмоции, классификация эмоциональной окраски текста может иметь различные уровни ранжирования – от бинарной (положительный или отрицательный) до множественной. Разные исследователи выделяют различное число базисных эмоций – от двух до десяти [1].

Важно понимать, что в лингвистике эмоции рассматриваются через призму их языкового выражения и эмоциональность в языке проявляется на всех его уровнях: фонетическом, морфологическом, лексическом и синтаксическом.

В тексте эмоциональность может создаваться комплексом языковых средств, включая метафоры, эпитеты, сравнения и другое. Важную роль играет и пунктуация, особенно восклицательные и вопросительные знаки, многоточие, тире. В письменной коммуникации, особенно в интернет-общении, для передачи эмоций часто используются дополнительные графические средства: эмодзи, смайлы, множественные знаки препинания [2]. Решение задач сентимент-анализа осложняется и такими особенностями представления текста, как метафоры, многозначность слов, сарказм и т.п. Выделение оценочных суждений является одним из методов анализа эмоциональной окраски текста. При этом подходе производится поиск и классификация языковых конструкций, выражающих субъективное мнение автора. Объектом тональности является объект или лицо, в отношении которого производится эмоциональное высказывание. Под субъектом понимают автора высказывания (текста, цитаты, прямой или косвенной речи), а под валентностью – эмоциональное отношение автора к объекту [3-4].

Для анализа тональности текста применяются методы компьютерной лингвистики такие как векторный анализ, который включает такие подходы, как наивный байесовский классификатор, случайный лес, рекуррентная нейронная сеть и другие. В последние годы активно разрабатываются подходы для классификации эмоций на основе методов машинного обучения и нейронных сетей. Все больше внимания уделяется также использованию больших языковых моделей на основе

трансформеров, которые способны моделировать контекст текста с высокой точностью и учитывать сложные зависимости между словами.

Анализ проблемы показал, что существующие методы, такие как наивный байесовский классификатор (LDA) [4] и метод опорных векторов (SVM) [5], имеют ограничения при работе со сложными контекстными зависимостями и не всегда способны учесть многообразие способов выражения эмоций в тексте. Поэтому больший интерес на данный момент представляют модели на основе трансформеров, способные эффективно обрабатывать контекстную информацию и выявлять сложные лингвистические конструкции.

В работе рассматривается решение задачи выявления и анализа эмоциональной окраски цифрового текста с помощью современных моделей машинного обучения. Проведен анализ существующих моделей и решений на их основе, таких как [6-9]: BERT, RoBERTa, DistilBERT, Encoder-Decoder, Multi-Head Attention, Self-Attention, Attention Mechanism. Каждая модель имеет свои преимущества и ограничения, и их выбор зависит от конкретной задачи, целей проекта и доступных ресурсов. Важно отметить, что ошибки на этапе выбора модели могут привести к серьезным проблемам, таким как дополнительные затраты на модификацию и исправление, неэффективность полученного решения и даже к полной неработоспособности. Одним из значительных преимуществ трансформеров перед RNN и LSTM является способность обрабатывать последовательности параллельно, а не поэтапно. Трансформеры позволяют значительно ускорить обучение, что особенно важно при обработке больших объемов данных.

В ходе исследования предметной области были рассмотрены различные аспекты анализа эмоциональной окраски текста и современные методы его реализации. Делается предположение, что для задач классификации эмоций может быть подходящим вариантом применение модели DistilBERT, которая является уменьшенной версией BERT, но при меньших вычислительных затратах сохраняет достаточный уровень точности. Эта модель идеально подходит для задач, где важна высокая скорость обработки и меньшие требования к памяти, но при этом точность остаётся на хорошем уровне. Предложен новый подход классификации эмоций с использованием в качестве базовой архитектуры модификацию модели BERT, дополнив её механизмами для многоуровневого анализа текста и интеграции локальных и глобальных результатов.

ЛИТЕРАТУРА

1. Изард, Кэррол Э. Психология эмоций / Кэррол Э. Изард ; [пер. с англ. В. Мисник, А. Татлыбаева]. – Москва [и др.] : Питер, 2006. – 460 с. : ил.

2. Иванова, И. В. Влияние эмодзи и эмотиконов на итоговую эмоциональную окраску текста при сентимент-анализе / И. В. Иванова, К.С. Пальмина // Научные исследования в современном мире. Теория и практика : Сборник избранных статей Всероссийской (национальной) научно-практической конференции, Санкт-Петербург, 10 мая 2021 года. – СПб: ЧНОУДПО ГНИИ «НАЦРАЗВИТИЕ», 2021. – С. 79-82.
3. Клековкина, М. В. Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики / М. В. Клековкина, Е. В. Котельников // Всероссийская конференция по электронным библиотекам: Digital libraries: Advanced Methods and Technologies, Digital Collections (Переславль-Залесский, 15-18 октября 2012 года). Переславль-Залесский: CEUR WS, 2012.
4. Пазельская, А. Метод определения эмоций в текстах на русском языке / А. Пазельская, А. Соловьев // The Intern. conf. on computational linguistics and intellectual technologies «Dialogue 2011». – М., 2011. – С. 510-522.
5. Турканов, Г. И. Классификатор Байеса для переменного количества признаков / Г. И. Турканов, Е. В. Щепин // Труды Московского физико-технического института (национального исследовательского университета). – 2016. – Т. 8, № 4(32). – С. 8-12.
6. Федотов, Д. В. О решении задачи классификации методом опорных векторов / Д. В. Федотов // Решетневские чтения. – 2013. – Т.2. – С. 77-79.
7. 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. – 2019. – С. 4171-4186.
8. Bidirectional Encoder Representations from Transformers. URL: https://tfhub.dev/tensorflow/bert_multi_cased_L-12_H-768_A-12/1.
9. Yinhan Liu et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach – arXiv: 1907.11692 [cs.CL], Jul. 2019.
10. Sanh, V., Debut, L., Chaumond, J. & Wolf, T. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. Preprint at URL: <https://doi.org/10.48550/arXiv.1910.01108> (2019).