

Павлюченко К.А.

Научный руководитель: Парамонов А.И., канд. техн. наук, доцент
Белорусский государственный университет информатики и радиоэлектроники

О РЕШЕНИИ ЗАДАЧИ ВЫЯВЛЕНИЯ ЭМОЦИОНАЛЬНОГО КОНТЕКСТА ДЛЯ ФРАГМЕНТОВ ТЕКСТА

***Аннотация.** Представлен метод анализа текста для выявления эмоционального контекста, который объединяет несколько компонент: классификатор эмоций, анализатор эмодзи, детектор сарказма и анализатор контекстных связей. Проведенная серия компьютерных экспериментов показала эффективность предложенного решения.*

***Ключевые слова:** тональность текста, сентимент-анализ, эмоции, контекст, машинное обучение, трансформеры, эмодзи, сарказм, обработка естественного языка.*

Рассматривается проблема определения тональности текстовых документов. Анализ эмоций, выраженных в текстах, сегодня является одной из приоритетных задач в области обработки естественно-языкового текста. Решение этой задачи поможет в принятии важных решений во многих отраслях экономики и сферах деятельности человека. Анализ проблемы позволил выявить тонкие места существующих методов анализа тональности (*Sentiment Analysis*), которые в полной мере неучиты-

вают такие нюансы эмоционального выражения, как сарказм, эмодзи и субкультурный контекст.

Разработанный метод анализирует текст с учетом различных аспектов эмоционального выражения: эмодзи, сарказм, специфические выражения и собственно классический анализ тональности слов. Каждый из этих аспектов эмоционального анализа вносит свой вклад в итоговый результат путем дополнения, уточнения эмоционального фона поэтапно. Метод основан на использовании трансформерной модели, которая реализует современную архитектуру энкодера RoBERTa[1]. Архитектура энкодера специально адаптирована и дополнительно обучена для решения задачи распознавания эмоций в текстах. Модель обучена распознавать семь эмоциональных состояний – шесть базовых и нейтральное. К положительным эмоциям относятся восхищение, доверие и принятие, а к отрицательным – печаль, гнев и отвращение.

Процесс классификации для каждого анализируемого текста проходит несколько этапов. Сначала модель разбивает текст на токены с учетом особенностей языка, применяя специализированный токенизатор, способный корректно обрабатывать морфологические особенности анализируемого языка. Затем токены трансформируются в семантические представления, которые отражают их значение и позицию в контексте. На следующем этапе модель анализирует контекстные связи между элементами текста, используя механизм многоголового самовнимания для выявления значимых взаимодействий между различными частями текста. В результате формируется скрытое представление текста, которое отражает его эмоциональную составляющую с учетом множества контекстуальных связей. Затем скрытое представление проецируется на выходной слой, генерируя вектор вероятностей эмоций. Для компенсации неравномерного распределения эмоций в обучающих данных применяются взвешенные функции потерь, которые придают больший вес недопредставленным классам. Для предотвращения переобучения и улучшения генерализации применяются методы регуляризации – *Dropout*[2] и *Weightdecay*[3], которые ограничивают сложность модели. Выходные вероятности корректируются для обеспечения надежной оценки уверенности модели, что особенно важно для последующей консолидации результатов с другими компонентами системы. Для оценки качества работы классификатора используется комплекс метрик. Общая точность отражает долю корректно классифицированных текстов среди всех анализируемых примеров. F1-мера для каждой категории эмоций позволяет оценить баланс между точностью и полнотой классификации, что особенно важно в условиях неравномерного распределения классов.

Для оценки качества вероятностных прогнозов и способности модели разделять различные классы используется *ROC-AUC*.

Анализатор эмодзи представляет собой специализированный компонент, который направлен на извлечение эмоционального фона за счет используемых в тексте символов эмодзи. После выделения эмодзи происходит их классификация по эмоциональной окраске. Каждый символ эмодзи сопоставляется с заранее определенным словарем, где каждому символу присвоена определенная эмоциональная категория.

Детектор сарказма основывается на распознавании характерных лингвистических паттернов, типичных для саркастических высказываний, которые могут существенно менять эмоциональный контекст сообщения. Ключевыми признаками сарказма выступают: конфликт между словами с негативной коннотацией и положительными эмодзи, а также использование определенных маркеров - особых фраз и конструкций, часто встречающихся в саркастических высказываниях. Большое внимание уделяется маркерам интенсификации – языковых средств, усиливающих эмоциональную окраску текста.

Механизм консолидации результатов анализа текста в разных «разрезах» и аспектах проявления эмоций представляет собой сложную логику взвешивания и объединения различных факторов, выявленных на предыдущих этапах анализа. Важной особенностью механизма является динамическая коррекция весов на основе уверенности отдельных классификаторов. К факторам, которые влияют на корректировку результатов, относятся длина анализируемого текста и наличие тех или иных рассматриваемых аспектов эмоции (эмодзи, сарказм и т.д.).

Проведенные эксперименты показали значительное улучшение точности определения эмоционального контекста по сравнению с традиционными методами анализа тональности. Общая точность классификации достигает 80.83%, что превосходит показатели базовых моделей, не учитывающих сарказм, эмодзи и другие контекстуальные факторы. Особенно заметное улучшение наблюдается при анализе сложных случаев, таких как тексты с саркастическим подтекстом или активным использованием эмодзи для модификации эмоционального содержания. Метод также показал улучшенную точность при работе с короткими сообщениями, характерными для современных платформ социальных медиа и мессенджеров. Это достигается благодаря способности системы извлекать максимум информации из ограниченного контекста путем комплексного анализа всех доступных эмоциональных маркеров.

СПИСОК ИСТОЧНИКОВ

1. Liu, Yinhan&Ott, Myle&Goyal, Naman&Du, Jingfei&Joshi, Mandar&Chen, Danqi&Levy, Omer&Lewis, Mike&Zettlemoyer, Luke&Stoyanov, Veselin. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. 10.48550/arXiv.1907.11692.
2. Ali, Shahwan. (2025). Dropout: An Effective Approach to Prevent Neural Networks from Overfitting. Asian Journal of Research in Computer Science. 18. 163-185. 10.9734/ajrcos/2025/v18i2569.
3. D'Angelo, Francesco & Andriushchenko, Maksym & Varre, Aditya & Flammarion, Nicolas. (2023). Why Do We Need Weight Decay in Modern Deep Learning?. 10.48550/arXiv.2310.04415v2 [cs.LG].