

УДК 004.896+316.6

МЕТОД ВЫЯВЛЕНИЕ ЭМОЦИОНАЛЬНОГО КОНТЕКСТА ФРАГМЕНТОВ ТЕКСТА

Павлюченко К.А., магистрант гр.356301

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Парамонов А.И. – канд. техн. наук, доцент

Аннотация. Представлен метод анализа эмоционального контекста текста, ориентированный на устранение недостатков существующих подходов к определению тональности. Предложено использование каскадной модели, состоящей из основного классификатора эмоций, анализатора эмпатии, детектора сарказма, контекстного анализатора и механизма консолидации. Модель основана на трансформере (RoBERTa или DistilBert), дополненном специальными модулями для распознавания тонких элементов эмоционального выражения. Экспериментально доказано значительное повышение точности анализа, особенно для коротких форматов общения (социальные сети, мессенджеры), содержащих эмоционально нагруженные элементы. Метод демонстрирует рост эффективности по сравнению с базовыми решениями, обеспечивая точное определение эмоционального состояния даже в сложных ситуациях.

Ключевые слова. Анализ эмоционального контекста, анализ тональности, сентимент-анализ, машинное обучение, трансформеры, эмпатия, сарказм, обработка естественного языка.

Введение.

Проблема анализа эмоций, передаваемых в тексте, становится одной из приоритетных задач в области автоматической обработки текста и востребована во многих отраслях. Особенно значима эта проблема в таких сферах, как маркетинг, социальные медиа и психология, где понимание эмоций способствует улучшению пользовательского опыта, повышению качества обслуживания и более точному прогнозированию поведения аудитории. Кроме того, все чаще про автоматическую обработку текста с целью выявления эмоций сообщений в социальных сетях говорят в контексте выявления фейковой информации и манипуляцией сознания [1].

Существующие методы анализа тональности зачастую не учитывают нюансы эмоционального выражения, такие как сарказм, использование эмпатии и культурный контекст. В данной статье представлен метод анализа эмоционального контекста, который направлен на решение этих вопросов.

Предлагается подход на основе многослойной обработки текстов, которая включает следующие компоненты (уровни): основной базовый классификатор эмоций; анализатор эмпатии; детектор сарказма; анализатор контекста и механизм консолидации. Каждый компонент учитывает определенный аспект сентимент анализа и вносит свой вклад в формирование итогового результата. Взаимодействие компонентов организовано по принципу последовательной модификации базового вектора эмоций с учетом дополнительных факторов, обнаруженных на каждом этапе анализа.

Основная часть.

В основе метода лежит трансформерная модель, базирующаяся на современной архитектуре энкодера (RoBERTa или DistilBert) [2, 3], специально адаптированная и дополнительно обученная для задачи распознавания эмоций в текстах. Выбор трансформерной модели обусловлен ее уже зарекомендовавшей себя эффективностью в задачах обработки естественного языка, когда требуется глубокое понимание контекста и семантики. Такие модели демонстрируют особую результативность при анализе текстов разной длины и структуры. Это критически важно для обработки текстов в цифровой коммуникации – сообщений, которые могут варьироваться от коротких реплик до развернутых высказываний.

Модель обучена распознавать семь эмоциональных состояний: шесть базовых и нейтральное. К положительным эмоциям отнесены восхищение, доверие и принятие. К отрицательным – печаль, гнев и отвращение. Нейтральное состояние маркирует отсутствие выраженной эмоциональной окраски. 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 и учитывает всю совокупность контекстуальных связей. Наконец, это скрытое представление проецируется на выходной слой, генерируя вектор вероятностей эмоций.

При разработке и обучении классификатора учитывались несколько важных особенностей. 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, такие как dropout и weight decay, которые ограничивают сложность модели. Выходные вероятности корректируются для обеспечения надежной оценки уверенности модели, что особенно важно для последующей интеграции результатов с другими компонентами системы.

При оценке качества классификатора используется комплекс метрик, обеспечивающих всестороннюю характеристику его производительности. 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. Матрица ошибок помогает выявить проблемные классы, между которыми чаще всего происходит смешение, и скорректировать модель соответствующим образом. Для оценки качества вероятностных прогнозов и способности модели разделять различные классы используется ROC-AUC [4]. Итоговый вектор вероятностей эмоций, полученный классификатором, служит базовым уровнем анализа, который затем модифицируется другими компонентами системы с учетом дополнительных факторов.

Вектор, полученный базовым классификатором, далее модифицируется с использованием компонента анализатора эмодзи. Анализатор эмодзи представляет собой специализированный компонент системы, направленный на извлечение дополнительной эмоциональной информации из 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 - положительная или отрицательная. На основе выделенных эмодзи и их классификации рассчитывается общий эмоциональный вклад эмодзи в контекст сообщения. Этот показатель учитывает не только количество положительных и отрицательных эмодзи, но и их относительную эмоциональную силу, позицию в тексте и взаимодействие с окружающими текстовыми элементами. Результатом работы анализатора является численная оценка эмоционального влияния эмодзи, которая в дальнейшем используется для корректировки базового вектора эмоций, полученного основным классификатором.

Следующим компонентом является детектор сарказма. Детектор сарказма представляет собой специализированный компонент системы, предназначенный для выявления саркастических высказываний, которые могут существенно менять эмоциональный контекст сообщения. Сарказм, как форма иронии, характеризуется противоречием между буквальным и подразумеваемым смыслом высказывания, что делает его распознавание сложной задачей для автоматического анализа. Разработанный детектор основывается на распознавании характерных лингвистических паттернов, типичных для саркастических высказываний. Один из ключевых признаков сарказма – это конфликт между словами с негативной коннотацией и положительными эмодзи. Например, фраза *"Какой замечательный день"* сопровождаемая эмодзи плача или злости, с высокой вероятностью является саркастической. Детектор выявляет такие несоответствия, анализируя лексический состав текста и сопоставляя его с эмоциональной окраской используемых эмодзи. Другим важным признаком сарказма является использование определенных маркеров – особых фраз и конструкций, часто встречающихся в саркастических высказываниях. К таким маркерам относятся выражения вроде *"ну да, конечно"*, *"как будто"*, *"якобы"*, *"ещё бы"*, сопровождаемые определенной пунктуацией или в определенном контексте. Детектор использует словарь таких маркеров и анализирует их присутствие и контекст использования в тексте. Анализ структуры предложений также играет значительную роль в выявлении сарказма. Характерные признаки включают сочетание вопросительных и восклицательных знаков, использование многоточий, преувеличений, повторений. Например, конструкции типа *"Неужели?!", "Да ладно???", "Вот это дааа..."* часто указывают на саркастический подтекст высказывания.

Все выявленные признаки сарказма объединяются с использованием взвешенной оценки, где вес каждого признака определен на основе его статистической значимости для определения сарказма. Результатом работы детектора является численная оценка вероятности наличия сарказма в тексте, которая в дальнейшем используется для корректировки базового вектора эмоций. При высокой вероятности сарказма происходит инверсия или существенная модификация исходного эмоционального вектора, отражающая истинный эмоциональный подтекст сообщения.

Контекстный анализатор является компонентом системы, который предназначен для анализа и учета широкого спектра контекстуальных факторов, влияющих на эмоциональную окраску текста. Этот компонент дополняет базовую классификацию эмоций, обогащая ее информацией о структурных, лексических и семантических особенностях анализируемого сообщения. Контекстный анализатор начинает работу с рассмотрения общего контекста сообщения, а далее переходит к рассмотрению отдельных предложений и их взаимосвязи. Важным аспектом является выявление логических и эмоциональных связей между предложениями – как они развивают, усиливают или изменяют эмоциональную линию текста.

Особое внимание уделяется выявлению маркеров интенсификации – языковых средств, усиливающих эмоциональную окраску текста. К таким маркерам относятся усилительные наречия (*"очень"*, *"крайне"*, *"невероятно"*), превосходные степени прилагательных, повторы. Контекстный

анализатор идентифицирует эти маркеры и оценивает их влияние на общую эмоциональную насыщенность текста. Не менее важным аспектом является анализ подтверждающих слов и конструкций, усиливающих уверенность в выражаемой эмоции. Такие слова как "несомненно", "безусловно", "точно", "однозначно" указывают на высокую степень уверенности в конкретной эмоции, что повышает надежность определения эмоционального контекста. Результатом работы контекстного анализатора является набор модификаторов, которые применяются к базовому вектору эмоций для его уточнения и обогащения. Эти модификаторы отражают влияние выявленных контекстуальных факторов на эмоциональную окраску текста и способствуют формированию более точной оценки эмоционального контекста сообщения.

На заключительном этапе реализуется механизм консолидации, который обеспечивает обобщение результатов работы всех предыдущих компонент системы в единую общую оценку эмоционального контекста анализируемого текста. Этот механизм реализует сложную логику взвешивания и объединения различных факторов, выявленных на предыдущих этапах анализа.

Первым шагом механизма является взвешенное объединение результатов отдельных компонент. Каждый компонент системы представляет свою оценку эмоционального содержания текста. Эти оценки объединяются с учетом предопределенных весов, отражающих относительную значимость каждого компонента для конкретного типа текста.

Важной особенностью механизма консолидации является динамическая коррекция весов на основе уверенности отдельных классификаторов. Если один из компонентов системы предоставляет результат с высокой степенью уверенности, его вес в итоговой оценке автоматически увеличивается. И наоборот, неуверенные оценки получают меньший вес. Этот подход позволяет системе адаптироваться к различным типам текстов и ситуациям, где определенные компоненты могут быть более или менее надежными. Механизм также включает набор правил корректировки для специфических случаев, требующих особого подхода. Например, при выявлении явного сарказма происходит инверсия эмоционального вектора – положительные эмоции, выраженные буквально, интерпретируются как отрицательные, и наоборот. Другим примером может быть корректировка при выявлении сильного влияния эмодзи на эмоциональную окраску текста, особенно если эмодзи противоречат буквальному содержанию сообщения.

После применения всех корректировок производится нормализация итогового вектора эмоций. Это необходимо для обеспечения совместимости результатов с последующими этапами обработки и для сохранения вероятностной интерпретации компонентов вектора. Нормализация гарантирует, что сумма всех компонент вектора равна единице, что позволяет рассматривать их как вероятности принадлежности текста к соответствующим эмоциональным категориям. Итоговый нормализованный вектор эмоций представляет собой комплексную оценку эмоционального контекста анализируемого текста с учетом всех выявленных факторов и особенностей.

Разработанный метод использует стратегию анализа на нескольких уровнях, от общего глобального анализа всего текста как единого целого до отдельных лексем. На этом уровне оценивается общая эмоциональная окраска сообщения, его тональность, преобладающие эмоциональные категории. Глобальный анализ позволяет получить общую картину эмоционального содержания текста и служит основой для дальнейшей детализации. Следующим уровнем является анализ на уровне отдельных предложений с учетом их индивидуальных особенностей.

Этот подход позволяет выявить эмоциональную динамику внутри текста, обнаружить эмоциональные пики и спады, идентифицировать предложения с контрастирующей эмоциональной окраской. Анализ на уровне предложений особенно важен для длинных текстов, где эмоциональная составляющая может значительно варьироваться от одной части текста к другой.

На первом этапе для каждого отдельного предложения и для текста в целом вычисляется исходный вектор эмоций:

$$E = [a, t, p, g, a, d], \quad (1)$$

где компоненты соответствуют вероятностям различных эмоциональных категорий: восхищение (a), доверие (t), принятие (p), печаль, (g), гнев (a), отвращение (d).

Этот вектор формируется основным классификатором эмоций и отражает базовую эмоциональную окраску текста без учета дополнительных факторов.

На следующем этапе применяется модификация вектора на основе выявленных особенностей:

$$E_m = E_0 \cdot M, \quad (2)$$

где M — матрица модификации, учитывающая влияние сарказма, эмодзи и других факторов, выявленных анализаторами. Эта матрица может усиливать или ослаблять определенные эмоциональные категории в зависимости от контекстуальных особенностей текста.

После модификации производится нормализация вектора:

$$E_n = \frac{E_m}{\sum E_m}. \quad (3)$$

Этот шаг необходим для сохранения вероятностной интерпретации компонент вектора и обеспечения их сопоставимости между различными текстами.

Комбинирование результатов на уровне текста и предложений выполняется на основе взвешенного среднего:

$$E_{final} = \alpha \cdot E_{text} + (1 - \alpha) \cdot \sum_{i=1}^n w_i E_{sentence_i}, \quad (4)$$

где α — вес глобального анализа, w_i — веса предложений на основе уверенности классификации для каждого предложения, n — количество предложений в тексте.

Такой подход позволяет учитывать как общую эмоциональную окраску текста, так и вклад отдельных предложений, придавая больший вес тем из них, которые классифицированы с высокой степенью уверенности.

Результаты и заключение.

Проведенные эксперименты показывают 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.

Следует отметить, что метод показал улучшенную точность при работе с короткими текстовыми 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. Сравнительный анализ с существующими методами определения эмоциональной окраски текста показал, что предложенный метод превосходит традиционные показатели в среднем на 15-20%.

Список использованных источников:

1. Парамонов, А. И. Фейковые новости и его влияние на информационное пространство: использование фейковых новостей в политических целях / А. И. Парамонов, М. В. Малышенко, В. В. Лысенко // Экономика устойчивого развития региона: инновации, финансовые аспекты, технологические драйверы развития в сфере туризма и гостеприимства : материалы XI Международной научно-практической конференции, Ялта, 26-29 марта 2024 г. / Крымский федеральный университет им. В. И. Вернадского ; отв. ред.: А. В. Олифинов. — Симферополь : ИТ «Ариал», 2024. — С. 155–158.
2. Liu, Yinhan et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach // arXiv preprint arXiv:1907.11692. — 2019. — DOI: 10.48550/arXiv.1907.11692.
3. Sanh, V., Debut, L., Chaumond, J. & Wolf, T. DistilBert, a distilled version of Bert: Smaller, Faster, Cheaper and Lighter // arXiv preprint arXiv:1910.01108. — 2019. — DOI: 10.48550/arXiv.1910.01108.
4. Fawcett, T. An introduction to ROC analysis // Pattern Recognition Letters. — Vol. 27, No. 8. — June 2006. — Pages 861–874. — DOI: 10.1016/j.patrec.2005.10.010.

UDC 004.896+316.6

METHODS FOR DETECTING THE TEXT EMOTIONAL CONTEXT

Paulichenka K.A., master's student

*Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Paramonov A.I. – Candidate of Technical Sciences, Associate Professor

Annotation. A method for analyzing the emotional context of text is presented, designed to address limitations of existing sentiment analysis techniques. A cascaded model is introduced, consisting of a base emotion classifier, emoji analyzer, sarcasm detector, context evaluator, and consolidation mechanism. Built upon transformer architectures (RoBERTa or DistilBert), enhanced with specialized modules for subtle emotional expressions, the model shows significant improvement in accuracy, particularly for short formats common in social media and messaging applications containing emotionally charged elements. Experimental results confirm increased efficiency compared to conventional methods, ensuring precise determination of emotional states even in complex situations.

Keywords. Emotional context analysis, Sentiment analysis, Tone analysis, Machine Learning (ML), Transformers, Emoji, Sarcasm, Natural Language Processing (NLP).