

ДЕТЕКЦИЯ САРКАЗМА В РУССКОЯЗЫЧНЫХ ТЕКСТАХ НА ОСНОВЕ МОДЕЛЕЙ С АРХИТЕКТУРОЙ BERT И ЛИНГВИСТИЧЕСКИХ ПРИЗНАКОВ

Мухин И. С.¹

Научный руководитель – канд. пед. наук, доцент Авксентьева Е. Ю.¹

¹Университет ИТМО
ilya.mukhinn@yandex.ru

Работа выполнена в рамках темы НИР №623106 «Автономные интеллектуальные системы»

Введение

Современные методы анализа эмоциональной тональности демонстрируют высокие показатели точности в классическом понимании решения задач данной области. Однако, остаются более узкие направления, которые всё ещё обладают ограничениями среди существующих решений. Одним из таких направлений является оценка тональности при наличии сарказма в анализируемом тексте. По оценкам исследований, в социальных медиа до ~12% сообщений могут содержать сарказм [1]. При этом исследования в данной области, связанные с русскоязычными текстами, довольно ограничены, и в наиболее близких работах, посвященных оценке сарказма или иронии, предоставляются значения F1 порядка 0.74–0.76 [2], иногда показатель доходит до 0.84 [3]. Таким образом, данное исследование призвано увеличить метрические показатели определения сарказма в русскоязычных текстовых данных, а также увеличить степень проработанности проблемы с помощью более детализированного подхода к определению того, что представляет из себя явление сарказма.

Основная часть

Саркастические выражения представляют собой конструкцию, где позитивная лексика скрывает негативный посыл, что требует глубокого подхода к анализу в отличие от бинарной оценки тональности. В письменной речи сарказм проявляется неоднозначно: от явных маркеров до скрытых смысловых конфликтов, основанных на контексте или культурном подтексте.

Цель работы — повышение качества детекции сарказма в русскоязычных текстах. Для этого предложены три оригинальных подхода:

Метод на основе лингвистических маркеров. К векторному представлению текста добавляются специфические признаки (CAPS, пунктуация, кавычки, эмодзи). Это интерпретируемое и экономичное решение эффективно выявляет сарказм с явными стилистическими индикаторами.

Метод переноса знаний через NLI. Перед решением основной задачи модель дообучается на задачах логического вывода (Russian SuperGLUE). Это позволяет улавливать неявные смысловые несоответствия по типу «похвала – упрёк», повышая качество распознавания скрытого сарказма.

Метод «эмоциональная неоднозначность». Экспериментальный подход, основанный на анализе конфликта тональностей: сарказм диагностируется по близости вероятностей позитивной и негативной окраски текста, что позволяет выявлять контраст между формой и содержанием.

В исследовании использовались трансформерные архитектуры (mBERT, RuBERT, RuRoBERTa) и комплекс корпусов данных, включающий тональные разметки и специализированные выборки сарказма.

Результаты и оптимальность решения

Сравнительный анализ показал, что предложенные методы эффективны в своих нишах. Подход на основе NLI-адаптации продемонстрировал наилучшие результаты,

подтверждая, что понимание логических противоречий является ключом к детекции сарказма. Метод лингвистических маркеров доказал устойчивость на случаях с явной стилистической окраской, оставаясь легковесным инструментом. Метод эмоциональной неоднозначности успешно подтвердил гипотезу о контрасте тональностей и может использоваться как диагностический признак в ансамблевых решениях.

Качественная оценка выявила, что комбинация методов позволяет охватить различные проявления сарказма. Наиболее сложными остаются контекстно-зависимые ситуации: диалоговый юмор, реплики, требующие знания предыстории, а также выражения, опирающиеся на культурный контекст.

Таким образом, предложено комплексное решение, сочетающее интерпретируемость лингвистических признаков, глубину семантического анализа через перенос знаний и диагностику через эмоциональный конфликт. Разработанные методы предлагают экономичные способы улучшения качества анализа за счёт использования предобученных моделей и дополнительных признаков, что подтверждается устойчивым приростом качества детекции.

Выводы

Были предложены три подхода к анализу сарказма, и было продемонстрировано, что на разных проявлениях сарказма они ведут себя предсказуемо. Однако, явления сарказма является более глубоким для анализа, основанного только на представленных подходами. Таким образом, для дальнейшего исследования — возможно, углубиться в усиление работы с контекстом, а также разделением юмора и сарказма.

Литература

1. Eke C. I., Norman A. A., & Shuib L. (2021) Context-Based Feature Technique for Sarcasm Identification in Benchmark Datasets Using Deep Learning and BERT Model. *IEEE Access*. 9, 48501–48518. <http://dx.doi.org/10.1109/ACCESS.2021.3068323>
2. Kosterin M., Paramonov I., & Lagutina N. (2023) Automatic Irony and Sarcasm Detection in Russian Sentences: Baseline Method. In *Proceedings of the 33rd Conference of Open Innovations Association*. Zilina, Slovakia, pp. 148–154. <http://dx.doi.org/10.23919/FRUCT58615.2023.10142992>
3. Костерин М. А., & Парамонов И. В. (2024) Применение глубоких нейронных сетей для автоматического определения иронии в русскоязычных текстах. *Моделирование и анализ информационных систем*. 31, 90–101. <https://doi.org/10.18255/1818-1015-2024-1-90-101>