

Анастасия Сергеевна Яновская, преподаватель
Минский государственный лингвистический
университет, Минск, Беларусь
э-почта: yanovska.nastya33@gmail.com

Nastassia Siarheyeuna Yanouskaya, Lecturer
Minsk State Linguistic University, Minsk,
Belarus
e-mail: yanovska.nastya33@gmail.com

АНАЛИЗ ТОНАЛЬНОСТИ В СОЦИАЛЬНЫХ МЕДИА: ВЫЗОВЫ ДЕТЕКЦИИ ЭМОЦИОНАЛЬНЫХ ОТТЕНКОВ В ЦИФРОВОЙ КОММУНИКАЦИИ

Проанализированы особенности анализа тональности комментариев в социальных медиа на примере видео инаугурации президента США. Рассмотрены проблемы автоматической детекции эмоциональной окраски текста, включая многозначность, сарказм и использование неформального языка.

Ключевые слова: анализ тональности; социальные медиа; обработка естественного языка; VADER, TextBlob.

SENTIMENT ANALYSIS IN SOCIAL MEDIA: CHALLENGES OF DETECTING EMOTIONAL NUANCES IN DIGITAL COMMUNICATION

The features of sentiment analysis of comments in social media are analyzed using the example of the U.S. presidential inauguration video. The challenges of automatic detection of emotional tone, including ambiguity, sarcasm, and informal language usage, are considered.

Key words: sentiment analysis; social media; natural language processing; VADER; TextBlob.

Анализ тональности как одна из задач компьютерной лингвистики представляет собой эффективный инструмент исследования эмоциональных оттенков, в том числе, в цифровой коммуникации, что способствует выявлению ключевых настроений пользователей и изучению социальных, политических и экономических процессов в условиях информационного общества.

Под термином *анализ тональности*, или *сентимент-анализ текста*, понимается процесс автоматического определения эмоциональной окраски высказывания, основанный на применении методов обработки естественного языка (NLP) и машинного обучения для анализа текстовых данных [1].

В настоящее время, как правило, выделяют несколько основных подходов к выявлению тональной направленности текста: *основанные на правилах, словарях и машинном обучении*. В рамках описываемого этапа исследования используется классический метод, основанный на словарях, или тональных лексиконах. Фактическим материалом исследования послужили комментарии пользователей, опубликованных на платформе YouTube. Объектом выражения мнений выбрано видео инаугурации президента США Дональда Трампа, что обусловлено значимостью данного события для политической повестки и высокой общественной активностью в комментариях к видео. Извлечение мнений происходило автоматическим способом при помощи разработанного программного обеспечения на языке программирования Python с использованием YouTube API.

В качестве первого инструмента для анализа тональности выбрана библиотека обработки естественного языка TextBlob. Текстовым данным, извлеченным из комментариев, присвоены метки «позитивный комментарий», «негативный комментарий», «нейтральный комментарий», что позволило получить базовую оценку эмоциональной окраски каждого текста. При вычислении процентного соотношения, были получены следующие данные:

- позитивных комментариев: 31.60%;
- негативных комментариев: 12.31%;
- нейтральных комментариев: 56.09%.

Высокий процент *нейтральных комментариев* может указывать на большую долю пользователей, которые не склонны выражать сильные эмоции в отношении данного политического события. Однако, обратившись к нескольким наглядным примерам комментариев, сделаем вывод о реальных причинах такого вывода.

Комментарий	Ожидаемая тональность	Возможная причина ошибки
<i>Please, American citizens, stand up for your freedom, for world freedom.</i>	Позитивная	Призывная форма может восприниматься как информативная, а не эмоциональная.
<i>God bless America with president Trump.</i>	Позитивная	Является распространенным выражением, которое может не восприниматься как ярко позитивное.
<i>American came back.</i>	Позитивная	Короткие фразы без выраженных эмоций сложны для автоматической классификации.
<i>A fool and his money are soon elected.</i>	Негативная	Игра слов, основанная на известной поговорке <i>A fool and his money are soon parted</i> 'Глупец и его деньги скоро расстанутся'. В данном варианте слово <i>parted</i> 'расставаться' заменено на <i>elected</i> 'избран', что превращает фразу в политический сарказм.

Основной причиной наличия погрешностей в анализе является недостаточная чувствительность метода к контексту и тонким эмоциональным оттенкам. Кроме того, сложности возникают при обработке саркастических и многозначных выражений, что снижает точность детекции эмоциональной окраски текста.

Для более точной обработки комментариев в социальных медиа был применен инструмент VADER (Valence Aware Dictionary and sEntiment Reasoner), который учитывает особенности неформальных текстов, таких как эмодзи, сокращения и специфические выражения, характерные для социальных сетей [2]. VADER, основанный на словарях с оценками эмоциональных окрашиваний, позволил более эффективно анализировать короткие и эмоционально насыщенные комментарии.

Результаты анализа показали, что 46,67% комментариев являются позитивными, 21,59% – негативными, а 31,74% – нейтральными. Высокая доля позитивных комментариев может быть связана с преобладанием аудитории, поддерживающей рассматриваемое политическое событие, а также с использованием патриотической риторики и хэштегов. Негативные комментарии включают критику политиков, сарказм и выражение недовольства, что иногда затрудняет автоматическую классификацию. Значительное количество нейтральных комментариев все так же свидетельствует о сложностях в определении тональности, особенно в случае политических аллюзий и контекстно-зависимых высказываний. Несмотря на эффективность метода VADER в анализе эмоционально выраженных сообщений, его ограничения проявляются в обработке сложных языковых конструкций и скрытой иронии.

Таким образом, проведенное исследование доказывает тот факт, что анализ тональности в социальных медиа сталкивается с рядом вызовов, таких как *многозначность слов, сарказм, ирония, использование неформального языка и эмодзи*, что делает задачу точной детекции эмоциональной окраски особенно сложной. В связи с этим, требуется использование усовершенствованных методов, которые способны эффективно справляться с такими сложностями и обеспечивать более точные результаты. В качестве перспективы исследования представляется целесообразным применение глубоких нейросетевых моделей, способных учитывать контекстные зависимости и латентные смыслы в текстах. Дополнительное внимание следует уделить созданию специализированных обучающих датасетов, ориентированных на политический дискурс, а также разработке гибридных подходов, объединяющих лексические и семантические методы анализа.

ЛИТЕРАТУРА

1. Pang B., Lee L. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2008. Vol. 2, No. 1–2. P. 1–135.
2. Hutto C. J., Gilbert E. E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Proceedings of the 8th International Conference on Weblogs and Social Media (ICWSM), 2014. P. 216–225.