

ДОСЛІДЖЕННЯ МЕТОДІВ АНАЛІЗУ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ТЕКСТІВ

Боков І.П., Петров К.Е.

e-mail: ihor.bokov@nure.ua, kostiantyn.petrov@nure.ua

Харківський національний університет радіоелектроніки, каф. ШІ
м. Харків, Україна

This work is devoted to the study of methods for determining the emotional coloring of texts. Traditional approaches, such as dictionary-based and rule-based methods, as well as modern machine learning techniques, including neural networks and transformer models were analyzed. Special attention is given to the challenges of sentiment analysis, such as sarcasm, irony, and contextual word meanings. The study highlights the advantages and limitations of different approaches and emphasizes the need for optimizing transformer models for resource-limited environments.

Визначення емоційного забарвлення тексту є однією з ключових задач обробки природної мови (NLP) та відіграє важливу роль у численних прикладних сферах, зокрема в маркетингу, соціології, психології, аналізі громадської думки та інформаційній безпеці [1]. Зростання текстових даних у соцмережах, новинах та на форумах підсилює потребу в автоматичному визначенні тональності. Системи аналізу тональності тексту дозволяють оперативно отримувати структуровану інформацію про емоційні настрої суспільства, прогнозувати реакцію на певні події, а також виявляти потенційні загрози чи деструктивний контент [2].

Попри значні успіхи в цій сфері, існуючі методи аналізу тональності мають ряд обмежень, які знижують їх ефективність. Зокрема, традиційні підходи часто не враховують контекстуального значення слів, що є критично важливим для точного розпізнавання емоційного забарвлення. Крім того, деякі алгоритми мають труднощі при аналізі багатозначних слів, сарказму, іронії та сленгових виразів. Тому є потреба у подальшому вдосконаленні методів аналізу тональності тексту, зокрема через поєднання кількох підходів та використання моделей глибокого навчання.

На сьогодні існує кілька основних підходів до визначення емоційного забарвлення текстів, кожен з яких має свої переваги та недоліки [3].

Одним із найбільш простих і поширених є словниковий метод, який базується на використанні попередньо створених лексичних ресурсів, де кожному слову присвоєно певну емоційну оцінку. Цей метод є ефективним для текстів, що містять явно виражені емоційно забарвлені слова, однак він не враховує контекстуального значення термінів.

Також популярними є методи, що засновані на правилах, які використовують певні лінгвістичні закономірності та синтаксичні особливості тексту для визначення його тональності. Такі методи

дозволяють більш точно визначати тональність речень у структурованих текстах, проте вони є складними у масштабуванні, оскільки кожна нова предметна область або мова потребує створення специфічних правил.

Методи машинного навчання, зокрема наївний байєсівський класифікатор, метод опорних векторів (SVM) та рекурентні нейронні мережі (RNN), показують значно кращі результати, ніж розглянуті вище підходи. Вони вміють адаптуватися до різних контекстів. Однак на якість їх роботи суттєво впливає обсяг та якість навчального датасету, а також збалансованість текстових даних.

Сучасні підходи до аналізу емоційного забарвлення тексту значною мірою орієнтовані на використання трансформерних моделей, таких як BERT (Bidirectional Encoder Representations from Transformers). Особливістю цієї моделі є можливість врахування контекстного значення слів у реченні та здатність розпізнавати складні мовні конструкції. Двонаправлена архітектура BERT забезпечує високу точність у класифікації текстів, але потребує значних ресурсів і великих розмічених датасетів.

Отже, трансформерні моделі забезпечують високу точність, але потребують значних ресурсів, тоді як традиційні методи менш ресурсоємні, але мають обмеження.

Головною метою дослідження є розробка власного комбінованого методу визначення емоційного забарвлення тексту, що поєднує лексичний аналіз, глибоке навчання на основі трансформерних моделей (BERT, RoBERTa) та модуль для розпізнавання сарказму й іронії. Оптимізація моделі включає зменшення кількості шарів та адаптацію до специфіки української мови. Запропонований метод тестувався на корпусах текстів із різними емоційними забарвленнями, демонструючи покращену точність порівняно з традиційними підходами. Використання комбінованого методу дозволяє ефективніше аналізувати контекст, розпізнавати складні мовні конструкції та працювати в умовах обмежених ресурсів, що робить його перспективним для застосування у сфері аналізу громадської думки в соціальних мережах та автоматизованих чат-ботах.

Список використаних джерел:

1. Бегунов А. Алгоритм автоматизованого визначення емоційного забарвлення текстових даних. Інженерія програмного забезпечення. 2011. №2(6). С. 93–97.
2. Bing L. Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. 1st edition. Cambridge : Cambridge University Press, 2015, 383 p.
3. Ghafoor Y., Jinping S., Calderon F. TERMS: textual emotion recognition in multidimensional space. The International Journal of Research on Intelligent Systems for Real Life Complex Problems. 2023. P. 2673–2693.