

УДК 004.8

СЕНТИМЕНТ-АНАЛИЗ КАК НАПРАВЛЕНИЕ ОБРАБОТКИ ЕСТЕСТВЕННОГО ЯЗЫКА

Вишневская А.А.

Студент

ФГБОУ ВО «Поволжский государственный университет телекоммуникаций и информатики»

Самара, Россия

Лиманова Н.И.

д.т.н., профессор, научный руководитель

ФГБОУ ВО «Поволжский государственный университет телекоммуникаций и информатики», ФГБОУ ВО «Самарский государственный технический университет»

Самара, Россия

Аннотация

В статье рассматривается sentiment-анализ как одно из ключевых направлений обработки естественного языка. Описываются основные понятия, задачи и области применения анализа тональности текстов. Также представлен краткий обзор современного состояния исследований в области NLP и роли нейросетевых методов в развитии автоматического анализа текстовой информации.

Ключевые слова: sentiment-анализ, обработка естественного языка, анализ тональности, машинное обучение, нейронные сети, NLP.

SENTIMENT ANALYSIS AS A FIELD OF NATURAL LANGUAGE PROCESSING

Vishnevskaya A.A.

Student

Volga Region State University of Telecommunications and Informatics

Samara, Russia

Limanova N.I.

Doctor of Technical Sciences, Professor, Scientific supervisor

Volga Region State University of Telecommunications and Informatics

Samara State Technical University

Samara, Russia

Abstract

The article examines sentiment analysis as one of the key areas of natural language processing. The main concepts, tasks, and application domains of sentiment analysis are described. A brief overview of the current state of research in NLP and the role of neural network methods in the development of automated text analysis is also presented.

Key words: sentiment analysis, natural language processing, opinion mining, machine learning, neural networks, NLP

Обработка естественного языка (Natural Language Processing, NLP) является одной из наиболее активно развивающихся областей искусственного интеллекта, направленной на создание методов и алгоритмов, позволяющих компьютерам анализировать, интерпретировать и генерировать текстовую информацию на естественных языках [1, 23]. В рамках данной области особое место занимает сентимент-анализ, или анализ тональности текстов, который ориентирован на выявление субъективной информации, эмоциональной окраски и оценочных суждений, содержащихся в текстах.

Рост объёмов текстовой информации в цифровой среде, связанный с развитием социальных сетей, онлайн-платформ, интернет-магазинов и сервисов отзывов, обусловил высокую практическую значимость автоматического анализа пользовательских текстов. Сентимент-анализ позволяет извлекать из неструктурированных текстов информацию об отношении пользователей к продуктам, услугам, событиям и явлениям, что делает его важным инструментом в задачах анализа общественного мнения, мониторинга репутации и интеллектуального анализа данных [2, 41].

Современные исследования в области сентимент-анализа опираются на достижения машинного обучения и нейросетевых технологий, что обеспечивает высокую точность классификации и возможность обработки больших массивов данных. В то же время, специфика естественного языка, его неоднозначность, контекстуальная зависимость и разнообразие форм выражения эмоций создают значительные научные и прикладные сложности, стимулирующие дальнейшее развитие данного направления [3].

Под сентимент-анализом принято понимать совокупность методов автоматического определения эмоциональной окраски текста, а также выявления субъективных оценок, выраженных автором по отношению к объекту высказывания [4]. В англоязычной литературе данный подход также известен под терминами *opinion mining* и *sentiment classification* [5, 3].

С формальной точки зрения задача сентимент-анализа может быть сведена к задаче классификации текстов, при которой каждому текстовому объекту сопоставляется одна из заранее определённых категорий, например: положительная, отрицательная или нейтральная тональность. В более сложных постановках возможна многоклассовая классификация, а также определение силы эмоциональной окраски или выявление отдельных аспектов оценки [6].

Сентимент-анализ основывается на предположении о том, что язык является носителем субъективных оценок и эмоций, которые могут быть формализованы и представлены в виде признаков, пригодных для автоматической обработки. Такие признаки могут включать лексические характеристики, синтаксические конструкции, семантические зависимости и контекстуальные особенности текста [7, 2].

В зависимости от уровня анализа различают сентимент-анализ на уровне документа, предложения и отдельных аспектов. Анализ на уровне документа предполагает определение общей тональности всего текста, в то время как аспектно-ориентированный подход позволяет выявлять отношение автора к конкретным характеристикам объекта, например, качеству обслуживания или цене товара [2].

Сентимент-анализ решает широкий спектр задач, связанных с автоматической интерпретацией субъективной информации. Одной из основных задач является классификация текстов по тональности, что активно применяется при анализе отзывов пользователей на товары и услуги, комментариев в социальных сетях и публикаций в средствах массовой информации [1].

В области маркетинга и управления репутацией сентимент-анализ используется для мониторинга общественного мнения и оценки восприятия брендов. Компании применяют данные методы для анализа клиентской удовлетворённости, выявления проблемных аспектов продуктов и прогнозирования потребительского поведения [5]. Анализ тональности позволяет обрабатывать большие объёмы данных в режиме реального времени, что особенно важно в условиях цифровой экономики.

В политических и социологических исследованиях сентимент-анализ применяется для изучения общественных настроений, анализа реакций на

политические события и оценки эффективности коммуникационных стратегий. В данной сфере автоматический анализ текстов дополняет традиционные методы социологических опросов, обеспечивая более оперативное получение информации.

Кроме того, sentiment-анализ используется в информационных системах поддержки принятия решений, интеллектуальных рекомендательных системах и системах мониторинга социальных медиа. В научных исследованиях методы анализа тональности применяются для изучения закономерностей языкового выражения эмоций и субъективных оценок.

Современное состояние исследований в области обработки естественного языка характеризуется активным внедрением методов глубокого обучения и нейросетевых архитектур. Традиционные подходы, основанные на правилах и статистических моделях, постепенно уступают место нейронным сетям, способным автоматически извлекать признаки из данных [3, 6].

Особую роль в развитии sentiment-анализа сыграли распределённые векторные представления слов, такие как Word2Vec и GloVe, позволяющие учитывать семантическую близость слов и контекст их употребления. Эти модели обеспечили качественный скачок в точности анализа текстов и стали стандартом в современных NLP-системах [5, 2].

Дальнейшее развитие получили рекуррентные нейронные сети и архитектуры с механизмами внимания, которые позволяют учитывать последовательную структуру текста и дальние зависимости между словами. На современном этапе значительное распространение получили трансформерные модели, такие как BERT и GPT, демонстрирующие высокую эффективность в широком спектре задач обработки языка.

Несмотря на значительный прогресс, задачи sentiment-анализа остаются актуальными и сложными. Проблемы иронии, сарказма, контекстной неоднозначности и языкового разнообразия требуют дальнейших исследований и разработки новых методов. В связи с этим sentiment-анализ продолжает оставаться одним из наиболее востребованных и динамично развивающихся направлений в области обработки естественного языка.

Библиографический список

1. Баева Е. А. Методы представления текстовых данных в задачах анализа тональности // Вестник ЮУрГУ. — 2020.
2. Liu B. Sentiment Analysis and Opinion Mining. — Morgan & Claypool Publishers, 2012.
3. Pang B., Lee L. Opinion Mining and Sentiment Analysis // Foundations and Trends in Information Retrieval. — 2008.
4. Баев А. А. Sentiment-анализ текстов: методы и приложения // Информационные технологии. — 2018. — № 4.
5. Mohammadi M., Hossein M. Sentiment Analysis: Concepts, Methods and Applications. — Springer, 2021.
6. Тихомиров Д. М., Коршунов А. А. Методы автоматического анализа тональности текстов. — М.: Физматлит, 2019.
7. Киссель А. А., Коваль М. В. Анализ тональности текстов на русском языке: подходы и методы. — СПб.: НИУ ИТМО, 2020.
8. Лиманова Н.И., Ковтун Д.С. ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА КАК ОСНОВА ЧАТ-БОТОВ // Бюллетень науки и практики. 2024. Т. 10. № 4. С. 426-429