



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ _____ «Информатика и системы управления»
КАФЕДРА _____ «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

«Обзор методов анализа тональности
естественно-языковых текстов»

Студент:	<u>ИУ7-53Б</u> (группа)	_____ (подпись, дата)	<u>М. Д. Маслова</u> (И. О. Фамилия)
Руководитель:		_____ (подпись, дата)	<u>А. А. Оленев</u> (И. О. Фамилия)

2022 г.

РЕФЕРАТ

Расчетно-пояснительная записка 10 с., 0 рис., 0 табл., 1 источн., 4 прил.
АНАЛИЗ ТОНАЛЬНОСТИ

СОДЕРЖАНИЕ

РЕФЕРАТ	2
ВВЕДЕНИЕ	4
1 Анализ предметной области	5
1.1 Основные определения	5
1.2 Актуальность задачи	5
1.3 Формализация задачи	6
1.4 Препроцессинг	6
2 Описание существующих решений	7
2.1 Методы основанные на лексике	7
2.2 Методы машинного обучения	7
2.2.1 Наивный Байес	7
2.2.2 Логическая регрессия	7
2.2.3 к ближайших соседей	7
2.2.4 что-то там про лес и деревья ;)	7
2.2.5 Нейронки	7
2.3 Гибридные	7
3 Классификация существующих решений	8
3.1 Технология метода	8
3.2 Уровни	8
3.3 От тематики???	8
3.4 Скорость	8
3.5 Данные/память	8
3.6 Точность	8
3.7 Время разработки	8
3.8 Язык???	8
3.9 По предварительной обработке данных	8
3.10 Влияние частей речи???	8
4 Заключение	9

ВВЕДЕНИЕ

1 Анализ предметной области

1.1 Основные определения

Sentiment analysis (SA) или анализ мнений извлекает и анализирует субъективную информацию из различных источников, таких как Интернет, социальные сети и другие источники, чтобы определить мнение людей с помощью обработки естественного языка (NLP), вычислительной лингвистики и анализа текста. Эта проанализированная информация дает представление о чувствах или отношении общества к определенным предметам, лицам или идеям и определяет контекстуальную полярность информации. [1]

Тональность

Анализ тональности

Предобработка

Корпус

Словарь

Позитивный/негативный/нейтральный

1.2 Актуальность задачи

Появление Интернета -> высказывание мнений -> возможность их анализа мнений/настроений. В этом и есть задача анализа. Здесь про текст. Сферы (продукты, политика, акции, социальное положение). Популярность -> много данных -> возможность получения точных результатов.

Но есть проблемы:

Накоплены большие объемы информации -> большая часть – текстовые данные. Несмотря на то, что текстовые данные содержат массу полезных знаний, они по своей природе являются неструктурированными, а существующие объемы данных не позволяют обработать тексты в ручную. При этом в отличие от традиционной обработки текста в анализе настроений незначительные вариации между двумя элементами текста существенно меняют смысл (например, добавление частицы "не"). Обработку русского языка затрудняет обильное использование носителями средств выразительности и переносных значений слов и раз.

Наличие большого количества сфер применений, необходимость решения описанных проблем приводит к созданию "каких-то там" методов.

1.3 Формализация задачи

Три класса текста. Задача в определении текста к одному из классов (без вероятности).

1.4 Препроцессинг

единый регистр

удаление пунктуации

лемматизация/стемминг

удаление стоп слов

удаление шума (хэш-тэгов, ссылок и тд)

2 Описание существующих решений

2.1 Методы основанные на лексике

Надо найти

2.2 Методы машинного обучения

2.2.1 Наивный Байес

2.2.2 Логическая регрессия

2.2.3 k ближайших соседей

2.2.4 что-то там про лес и деревья ;)

2.2.5 Нейронки

2.3 Гибридные

Общее описание

3 Классификация существующих решений

3.1 Технология метода

machine learning

lexicon approach

hybrid

3.2 Уровни

Здесь необходимо пояснить за семантические связи. "Еда вкусная, но обслуживание так себе". В целом – скорее всего нейтральный, но по аспектам: о еде: положительно, обобслуживании: отрицательно.

document

sentence

approach

3.3 От тематики???

3.4 Скорость

3.5 Данные/память

3.6 Точность

3.7 Время разработки

3.8 Язык???

3.9 По предварительной обработке данных

3.10 Влияние частей речи???

4 Заключение

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Khalil E., El Houby E., Mohamed H.* Sentiment Analysis Tasks and Approaches // International Journal of Computer Science and Information Security. — 2021. — Т. 19. — URL: https://www.researchgate.net/publication/356068600_Sentiment_Analysis_Tasks_and_Approaches (дата обращения: 15.12.2021).