

РЕФЕРАТ

Выпускная квалификационная работа содержит 11 страниц, 2 рисунка, 1 таблицу, 3 использованных источника.

КЛЮЧЕВОЕ СЛОВО 1, КЛЮЧЕВОЕ СЛОВО 2, ...

Краткое описание содержания работы.

СОДЕРЖАНИЕ

Введение	4
1 Теоретическая часть	6
1.1 Сентимент-анализ текстов	6
1.1.1 Методы, основанные на правилах и словарях	7
2 Практическая часть	8
Заключение	9
Список использованных источников	10
Приложение	11

ВВЕДЕНИЕ

Классификация текстов — невероятно популярная задача. Мы пользуемся текстовыми классификаторами в почтовом клиенте: он классифицирует письма и фильтрует спам. Другие приложения включают классификацию документов, обзоров и так далее.

Обычно классификация текстов используется не как самостоятельная задача, а является частью более крупного пайплайна. Например, голосовой помощник классифицирует ваше высказывание, чтобы понять, что вы хотите (например, установить будильник, заказать такси или просто поболтать) и передать ваше сообщение другой модели в зависимости от решения классификатора. Другой пример — поисковый движок: он может использовать классификаторы для определения языка запроса, чтобы предсказать его тип (например, информационный, навигационный, транзакционный) и понять хотите ли вы увидеть картинки или видео помимо документов и прочего.

Моя работа посвящена одному из приложений классификации — автоматическому определению эмоциональной оценки русскоязычных текстов. Главная особенность заключается в том, что предсказание базируется на эмоциональных моделях, предложенных Робертом Плутчиком и Полом Экманом.

Поскольку большинство датасетов для классификации содержат только одну правильную метку. А у нас многоклассовая классификация.

ОСНОВНАЯ ЧАСТЬ

ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

В этой части мы подробнее раскроем теорию.

1.1. СЕНТИМЕНТ-АНАЛИЗ ТЕКСТОВ

ОПРЕДЕЛЕНИЕ 1. Классификация текста — использование математических моделей для определения принадлежности текста, на основании его содержания, к конкретному классу из определенного множества.

ОПРЕДЕЛЕНИЕ 2. Мультиклассификатор текстов — алгоритм или обученная модель, предсказывающие принадлежность каждого входного текста к одному или нескольким классам, множество классов определено заранее.

ОПРЕДЕЛЕНИЕ 3. Сентимент-анализ (анализ тональности) — автоматическое выявление в текстах эмоционально окрашенной лексики и эмоциональной оценки авторов.

ОПРЕДЕЛЕНИЕ 4. Сентимент — совокупность чувств и взглядов, как основа для действия или суждения; общая эмоциональная установка.

Сентимент-анализ, как направление компьютерной лингвистики, стал очень популярен последние десятилетие. С появлением больших данных насыщенных эмоциональной составляющей возникла потребность в их обработке. Компании начали устраивать соревнования с внушительными призовыми фондами, а исследователи искать лучшую архитектуру для классификации этих данных. Так в открытом доступе появились большие размеченные датасеты с отзывами, данными из соцсетей и новостями.

Сам термин «сентимент» означает совокупность чувств и взглядов, как основа для действия или суждения; общая эмоциональная установка. Целью сентимент-анализа является выделение этих тональных компонент из текста. Рассмотрим его применение на разных уровнях.

Пусть есть целый документ, тогда, как правило, в нем можно выделить один субъект и один объект, а так же сентимент. Ярким примером такого уровня является отзыв. Здесь автор выступает в качестве субъекта, а предмет отзыва — в качестве объекта. Это уровень документа.

Если документ более сложный, то можно рассматривать его на уровне предложений. В результате можно определить эмоциональную окраску всего документа или предложений по отдельности. Зависит от поставленной задачи.

Также анализ бывает на уровне аспектов. Смысл его в том, что эмоциональная установка определяется не для конкретно объекта, а для его отдельных составляющих — аспектов. Например, для объекта «компьютер» можно выделить аспекты «производительность», «дизайн», «сборка» т.д., другими словами, к аспектам относится все то, к чему могут быть выражены эмоции. Данная задача очень востребована, потому что позволяет точнее определять отношение автора к объекту, а для некоторых задач это очень важный критерий.

Последний вид анализа самый сложный и проводится на уровне именованных сущностей (Named Entities). Именованная сущность — это абстрактный или физический объект, который может быть обозначен собственным именем. Сама по себе задача извлечения именованных сущностей (Named-entity recognition) не из простых, а вкупе с сентимент-анализом становится действительно комплексным решением.

Исследователи уже достигли внушительных результатов в сентимент-анализе английского

Во многом благодаря выдающимся работам ++, которые помогли

1.1.1. МЕТОДЫ, ОСНОВАННЫЕ НА ПРАВИЛАХ И СЛОВАРЯХ

ПРАКТИЧЕСКАЯ ЧАСТЬ

ЗАКЛЮЧЕНИЕ

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Ильин В. А., Позняк Э. Г. Основы математического анализа: в 2-х ч. Часть II: учеб.: для вузов. — 4-е изд. — М.: ФИЗМАТЛИТ, 2002. — 464 с. — (Курс высшей математики и математической физики).
2. Semina T. A. Sentiment analysis: Modern approaches and existing problems. // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 6: Языкознание. Реферативный журнал — 2020. — № 4. — С. 47–53
3. Абгарян К. А. К теории балок минимального веса. // М.: Машгиз, 1962. — вып. 8 — С. 136–151.

ПРИЛОЖЕНИЕ
