

# README

---

## 项目简介

---

文本情感分析又称意见挖掘，是对包含用户观点、喜好、情感等主观性文本进行挖掘、分析及判别。它是一个多学科交叉的研究领域，涉及概率论、数据统计分析、计算机语言学、自然语言处理、机器学习、信息检索、本体学 (Ontology) 等多个学科及其相关技术。

目前，情感分类大致涌现出两种研究思路：基于情感知识和基于特征。前者主要是基于已有的情感词典或情感知识库对文本中带有情感或极性的词(或词语单元)进行加权求和，而后者主要是对文本提取具有类别表征意义的特征，再基于这些特征使用机器学习算法进行分类。

本项目运用情感词典与机器学习两种方法分别进行文本情感分析，并提供结果对比。

## 项目构建方法

---

### 环境准备

- Windows/MacOS/Linux
- Python3.6
- PyCharm or other IDEs

### 获取项目

- get the code from gitlab/github

```
git clone git@github.com:Charon0622/Software-Engineering-Course-Design.git
```

### 导入项目

Open the file named "Chinese-emotion-analysis" with IDE

## 项目运行方法

---

### 本地运行

```
cd [project folder]
python3 manager.py runserver [port]
```

### 直接访问

<http://115.28.245.233:8080>

## 项目基本功能

---

机器学习方法的接口，接受一个中文文本， 可得到一个正向情感极性的概率和负向情感的概率。

基于情感词典的方法的接口， 输入一段中文文本， 可得到文本的情感极性分值 。

可对 篇章级、段落级、句子级 的中文文本进行情感极性判断。

基于情感词典的方法的接口， 可以使用不同的情感词典对中文文本进行情感分析。

基于机器学习的方法的接口， 可以导入一个训练好的模型来对中文文本进行情感分析。

分析算法以及修复程序中的bug。

## 代码结构说明

---

Chinese\_Emotion\_Analysis

settings.py web的总配置文件

url.py web的路由配置

Emotion\_Manager

CEA\_LIB NLP分析库

pkl\_data 处理之后的数据

raw\_data 未处理之前的数据

chinese\_emotion\_analysis.py NLP方法的接口

classifier.pkl 训练好的分类器模型

Modules 词典方法分析库

res字典存放处

main.py 接口

migrations 数据连接层

static 静态资源文件

templates 网页模板

models.py 模型构建

views.py web逻辑