

老师您好！下面是此项目文件的相关信息：

## 一、python 推荐版本：3.10

## 二、所需库：

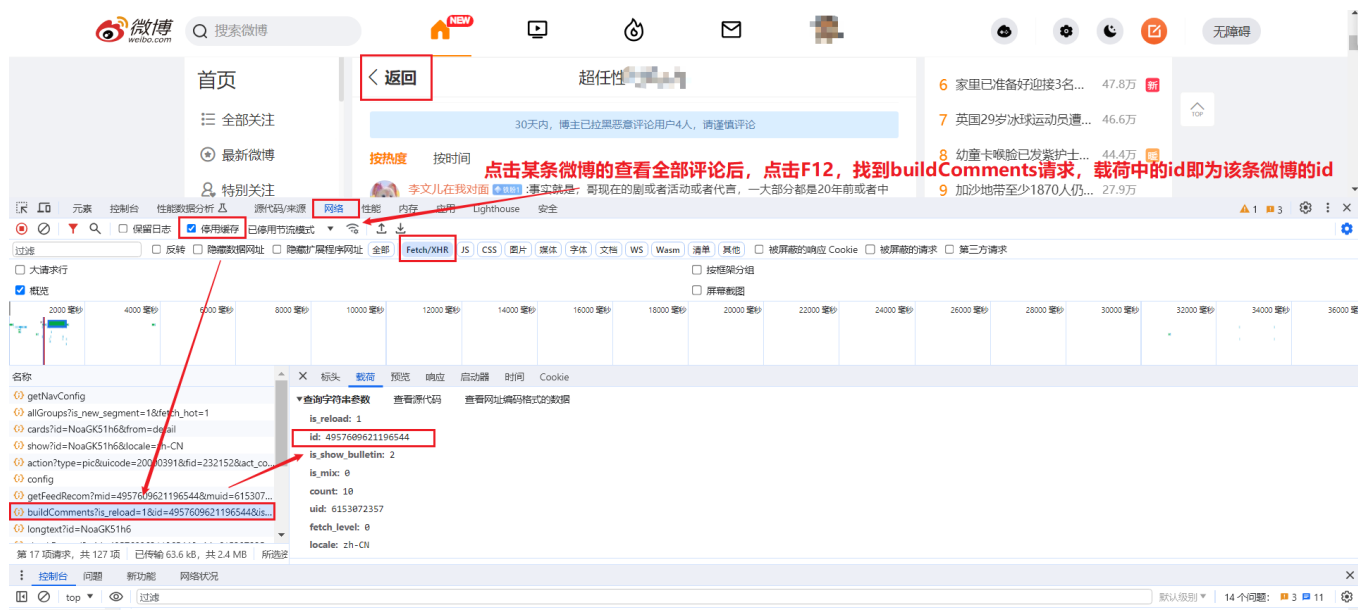
pandas  
matplotlib  
numpy  
json  
csv  
requests  
time  
fake\_useragent  
seaborn  
joypy  
openpyxl  
snownlp  
Pillow (PIL)  
base64  
io  
jieba  
wordcloud  
imageio  
os

## 三、项目介绍

本项目基于 snownlp 库，可自动爬取目标微博评论区的评论，生成相关数据图、数据文件和情感分析报告以及标签化评论区

## 四、使用方法

### 1. 查看目标微博的 id



在 examples 文件夹中有各示例帖子的 id

📁 梅西道歉	2024/3/15 12:28	文件夹	
📁 女神节	2024/3/15 11:43	文件夹	
📁 热辣滚烫	2024/3/15 11:49	文件夹	
📁 僧丫事件	2024/3/15 12:08	文件夹	
📁 鼠头鸭脖	2024/3/15 12:12	文件夹	
📄 各帖子id.txt	2024/3/15 10:03	文本文档	1 KB

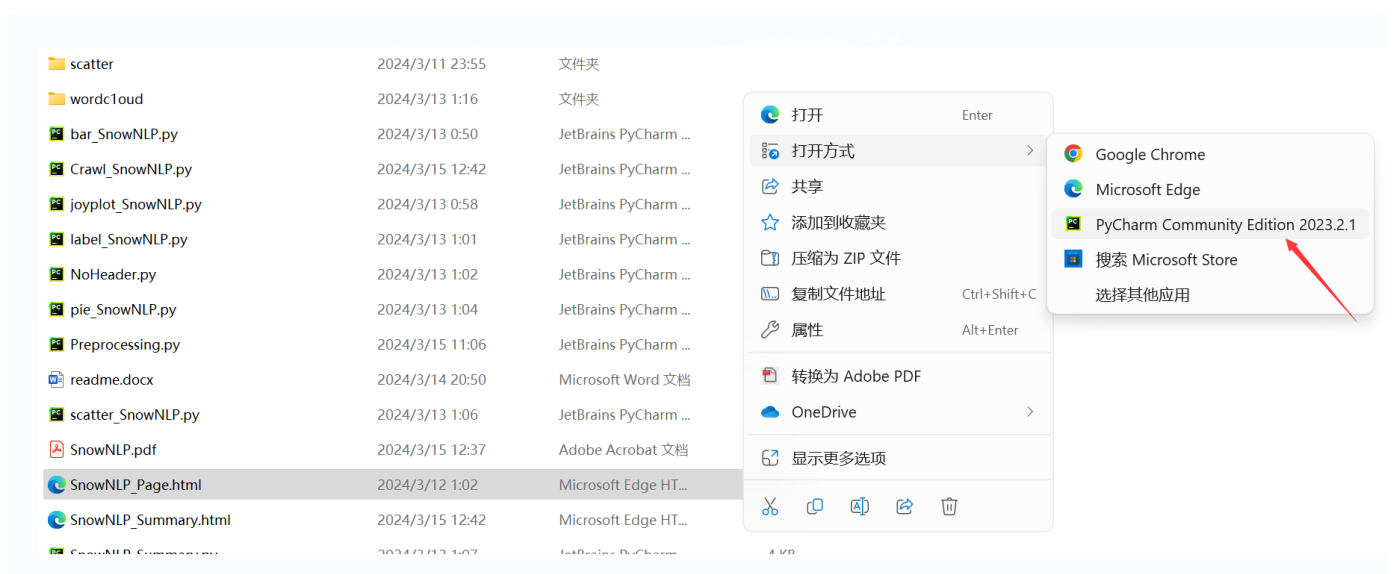
## 2. 将 Crawl\_SnowNLP.py 中79行的相应位置改为目标微博的 id

```
76
77 if __name__ == '__main__':
78
79     weibo_ids = ["4965074095576055"] # 目标微博的id
80     filename = "files/data.csv"
81     if os.path.exists(filename):
82         os.remove(filename)
83     for weibo_id in weibo_ids:
84         print("爬取开始")
85
```

## 3. 运行 Startup\_SnowNLP.py

## 4. 直接打开 SnowNLP\_Summary.html 查看分析报告

## 5. 选择 pycharm 打开 SnowNLP\_Page.html，在 pycharm 中用浏览器打开，生成标签化评论区（直接点开将无法正常载入评论）



The screenshot shows a file explorer window with a list of files and folders. A context menu is open over the file 'SnowNLP\_Page.html'. The menu options include '打开' (Open), '打开方式' (Open with), '共享' (Share), '添加到收藏夹' (Add to favorites), '压缩为 ZIP 文件' (Compress to ZIP file), '复制文件地址' (Copy file address), '属性' (Properties), '转换为 Adobe PDF' (Convert to Adobe PDF), 'OneDrive', and '显示更多选项' (Show more options). The '打开方式' option is expanded, showing a list of applications: Google Chrome, Microsoft Edge, PyCharm Community Edition 2023.2.1, and '搜索 Microsoft Store' (Search Microsoft Store). A red arrow points to 'PyCharm Community Edition 2023.2.1'.

```
<!DOCTYPE html>
<html lang="zh">
<head>
  <meta charset="UTF-8">
  <meta name="viewport" content="width=device-width, initial-scale=1.0">
  <title>Labels Based on SnowNLP</title>
  <style>
    body {
      font-family: Arial, sans-serif;
      margin: 0;
      padding: 0;
      background-color: #f0f0f0;
    }
  </style>
</head>
```

# 基于SnowNLP的微博评论情感分析

0-0.2

0.2-0.4

0.4-0.6

0.6-0.8

0.8-1.0

不懂事，但懂得骗中国人的钱

你咋不说说主办方的问题呢？早就被扒得赤裸裸的了还在拿梅西说事[doge]

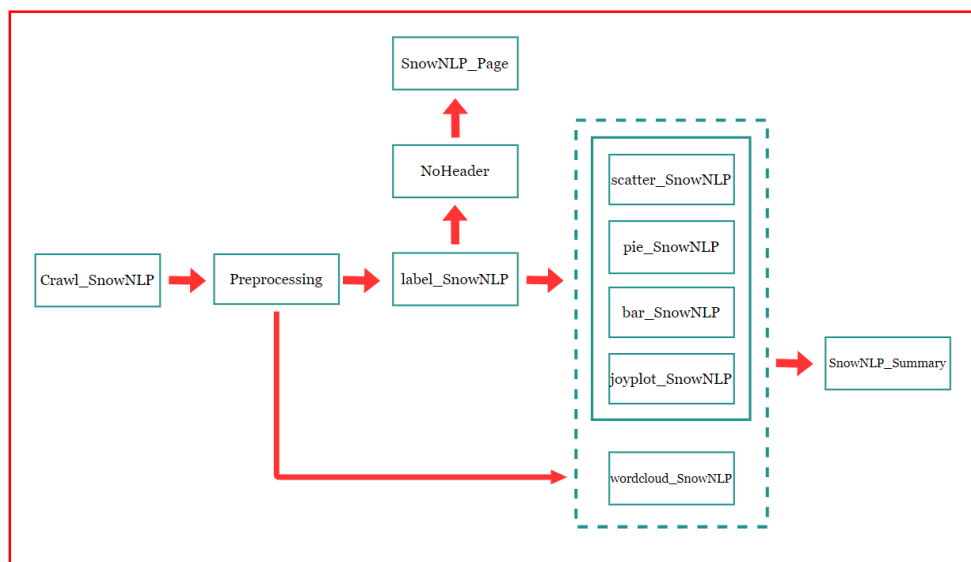
就一踢足球的，有什么屌的。不看球会死吗？[吃瓜]

你以为他是爱梅西？这是收了定金 现在得退钱 说不定还要赔呢

说得好！梅西三月行取消，并非是官方“站队”，只是怕喷子太多“不稳定”！

## 五、程序说明

### Startup\_SnowNLP



Crawl\_SnowNLP.py

爬取目标微博评论区，并保存为csv文件（files/data.csv）

Preprocessing.py

对csv文件进行预处理，包括添加列标题，删去空行，保存为xlsx文件（files/comments.xlsx）

label\_SnowNLP.py

使用 SnowNLP 对评论逐条进行情感打分，并按得分将其均匀分为五组，相关 xlsx 文件均存入 labels 文件夹

NoHeader.py

去除comments\_score.xlsx内数据的列名，为生成标签化评论区做准备

scatter\_SnowNLP.py

根据分析结果，绘制情感得分散点图，呈现得分的分布样态，图片存入scatter文件夹

pie\_SnowNLP.py

根据分析结果，绘制情感得分散点图，呈现得分的分布样态，图片存入pie文件夹

bar\_SnowNLP  
根据分组结果，绘制各组数量的条形图，图片存入bar文件夹

joyplot\_SnowNLP  
根据各组内部的情感得分分布情况，绘制山脊图，图片存入joyplot文件夹







wordcloud\_SnowNLP.py  
根据所有评论的文本，绘制词云图，图片存入wordcloud文件夹

SnowNLP\_Summary.py  
根据以上步骤绘制的图表与数据文件，生成一份综合报告HTML文件

SnowNLP\_Page.html  
在 pycharm 中用浏览器打开，生成标签化评论区

## 六、示例文件

在 examples 文件夹中有各事件的数据结果，分析报告

 梅西道歉	2024/3/15 12:28	文件夹	
 女神节	2024/3/15 11:43	文件夹	
 热辣滚烫	2024/3/15 11:49	文件夹	
 僧丫事件	2024/3/15 12:08	文件夹	
 鼠头鸭脖	2024/3/15 12:12	文件夹	
 各帖子id.txt	2024/3/15 10:03	文本文档	1 KB