

使用说明

数据集部分

数据集在data目录下，里面包含了测试集数据[tmp_test_data.csv](#)，大约有5000条评论数据，共11款手机，便于演示。同时也包含了完整的数据集[JDComment_data](#)，大约2，包含进60款手机的评论数据。[result.csv](#)是采用完整数据集计算出的各个手机的评论得分，可以用做功能演示。

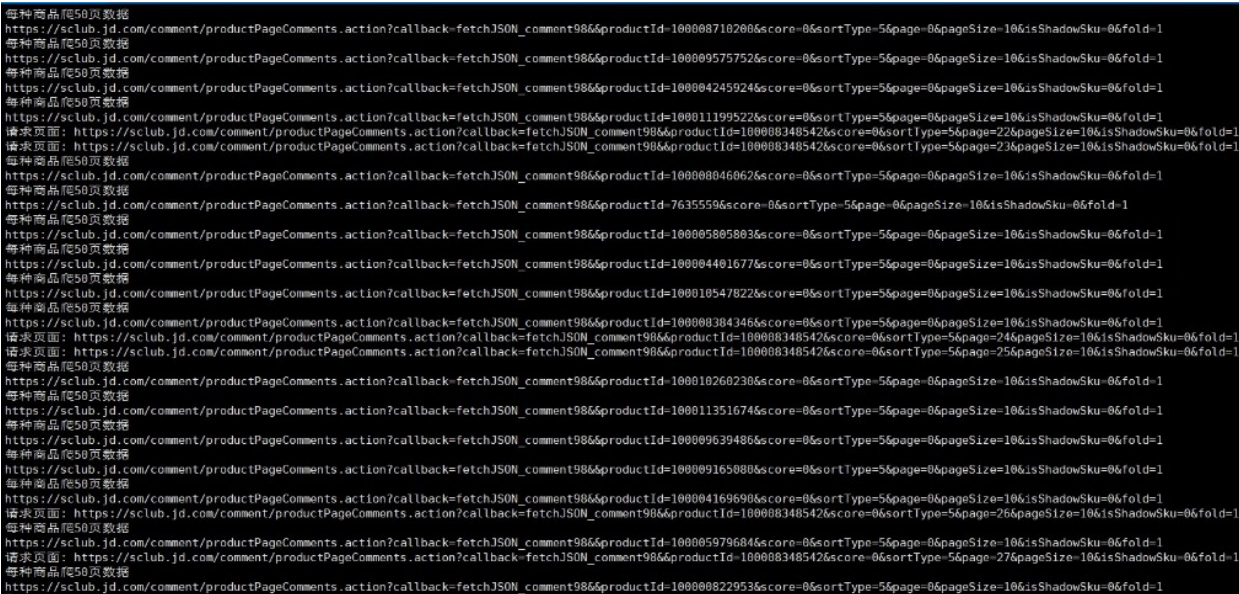
评论数据采集

除了已经采集好的数据集，也可以通过脚本SpiderScript重新爬取

在安装好Python，以及配置好pip或conda环境之后，在[当前位置打开cmd控制台](#)(windows)或者在终端输入(Linux)以下语句执行脚本

```
python SpiderScript.py
```

正常执行界面如下图所示，采集完成之后保存到路径data/JDComment_data中



```
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008710200&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100009575752&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100004245924&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100011199522&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
请求页面: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008348542&score=0&sortType=56page=226pageSize=10&isShadowSku=0&fold=1
请求页面: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008348542&score=0&sortType=56page=236pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008046062&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=76355596&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100005805803&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100004401677&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100010547822&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008384346&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
请求页面: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008348542&score=0&sortType=56page=246pageSize=10&isShadowSku=0&fold=1
请求页面: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008348542&score=0&sortType=56page=256pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100010260230&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100011351074&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100009639486&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100009165080&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100004169690&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
请求页面: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008348542&score=0&sortType=56page=266pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100005979684&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
请求页面: https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=100008348542&score=0&sortType=56page=276pageSize=10&isShadowSku=0&fold=1
每种商品爬50页数据
https://club.jd.com/comment/productPageComments.action?callback=fetchJSON_comment986&productID=10000822953&score=0&sortType=56page=06pageSize=10&isShadowSku=0&fold=1
爬取完成，数据保存到data/JDComment_data中
```

评论数据情感分析并计算得分

tmp.py和comment_analysis_process都是计算情感分析的python源码，区别只是在于文件格式不同而已，可以根据不同的环境采用不同的脚本。

tmp.py运行方式为在控制台输入 `python tmp.py` (windows)

comment_analysis_process可以用jupyter notebook或其他ipython IDE打开执行

页面测试

双击打开目录下的index.html，上传data目录下的result数据集进行测试即可，只能在本地演示。

版本控制

为了便于管理和维护，我已将项目提交到https://github.com/YuleZhang/NLP_Analysis_IDcomment，可以自行下载查阅。

部分参考

[python实现多线程爬虫](#)

[MySQL的python连接](#)

[用python实现文本情感分析](#)