

未名 bbs 评论区情绪评估与心理疏导资源分配

Github 链接:https://github.com/hcx1999/BBS_NLP

1. 问题需求分析

1.1 研究背景

大学生的心理健康问题越来越受到关注，尤其是在学业压力、就业压力和生活压力交织的情况下。未名 BBS 作为一个大学生广泛使用的交流平台，包含了大量的评论和帖子，这些内容可以反映出学生的情绪状态。通过对这些评论进行情绪评估，可以帮助学校更好地了解学生的心理状态，并及时分配心理疏导资源。NLP 可以进行实时情绪分析，即时洞察学生的情绪状态。这对于在出现困境或危机时及时干预至关重要。

1.2 研究目标

本研究的目的在于，分析未名 BBS 评论区的情绪状态，以了解大学生在不同时间段内的压力变化。研究大学生在临近毕业季等关键时间点的情绪变化情况。并且提供数据支持，帮助学校在高压力时期更有效地分配心理疏导资源。研究者可以在不侵入受试者的情况下采集到真实情境中的数据，为心理危机状态的预测、心理异常用户的识别、心理健康问题的干预提供多元化的证据，促进学生的身心健康发展，助力高校智能化心理健康平台建设。

1.3 研究范围

《2022 年大学生心理健康状况报告》显示，调查中，39%的心声认为没有挚友；24%的学生感到孤独；43%的学生更希望自己成为沟通的对象儿不是沟通的干脆发起者。与此同时，由于个体间的正常交往不够，又简单引起猜疑、嫉妒等，极不利于学生的心理健康发展。调查显示，在人际交往中，在遇到不被人理解的时候，有 75%的学生会感到难过，会产生郁闷心理。一些严峻的人际交往问题困扰着少数同学，假如不解决这些问题，长期压抑的积累将会导致严重的后果。面对自己存在的心理问题，绝大部分的学生能够客观的正视并努力寻求方法解决。因此，我们选取的研究对象是未名 BBS 评论区的大学生用户。选取的数据范围为：未名 BBS 评论区内的所有公开评论和帖子；时间范围为，为期一年的评论数据，重点关注临近考试期间、毕业季等。

1.4 关键术语定义

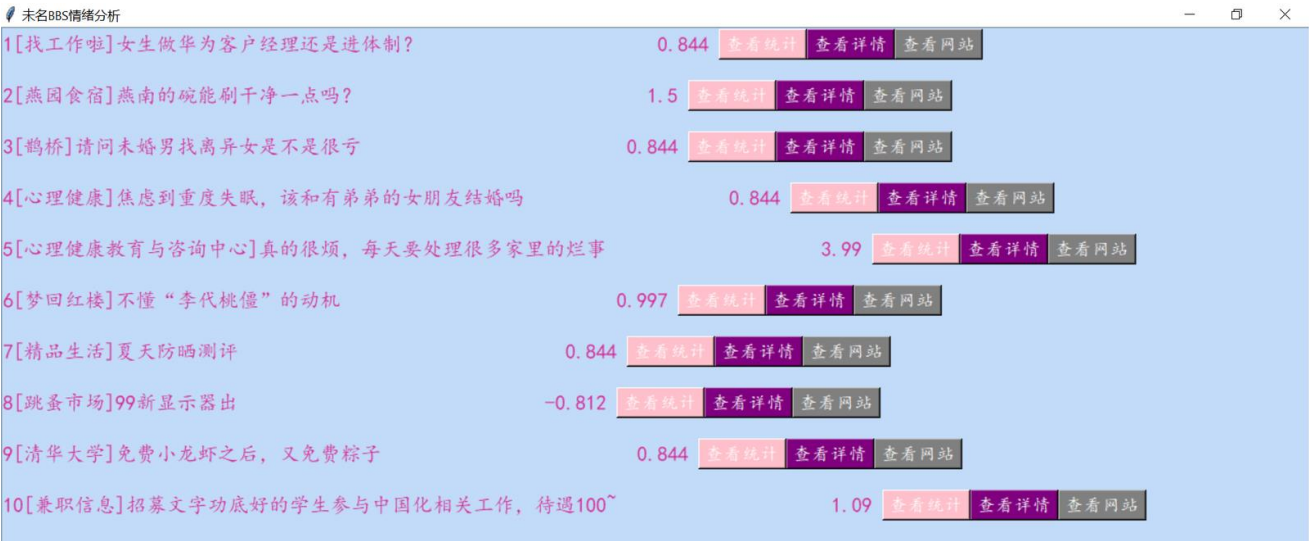
情绪评估：通过自然语言处理工具（如 SnowNLP）对评论进行情绪分析，获得情绪评分。

压力指标：基于情绪评分，确定评论中反映的压力程度。

心理疏导资源：学校提供的心理咨询、心理辅导、减压活动等资源。

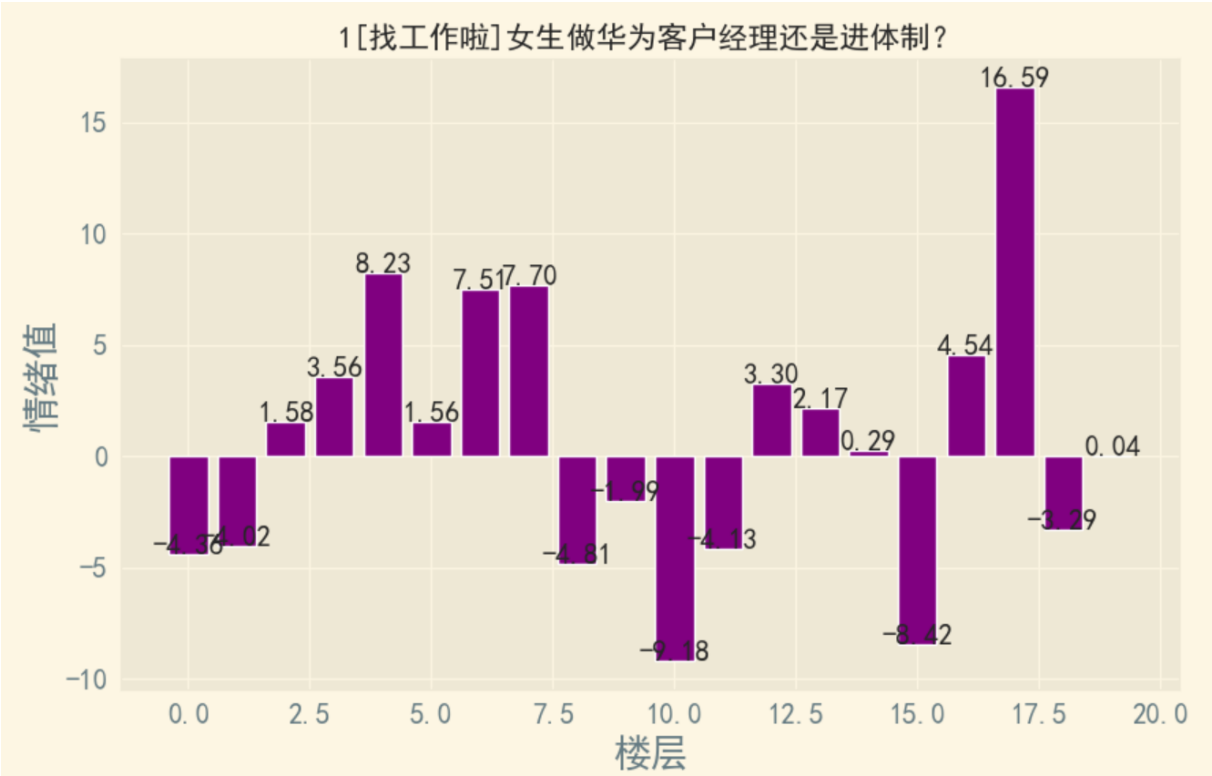
2. 界面设计

2.1 整体界面：



2.2 具体功能截图

提供了查看统计、查看详情、查看原网站等三个功能，具体工作截图：



4	A	B	C
1	楼层	内容	情绪
2	1楼	人文社科大三，保研悬出国难，只能工作的话能做什么准备呢？绩点绝对值还行（3.7+）但排名中游，只能拿到推免保研无缘top2，需要降级比较严重，且感觉现在来不及准备保研了；出国负担较为沉重，最多就是港一年或者欧陆无学费的，其他的真不行了，最多只有30w预算，这已经掏空家底了。因为家境实在窘迫（从出国也能看出来），加上感觉读人文社科的作用也不大，目前想着直接就业，有下面几个打算，还请大家提一些建议：	-4.36369
3	2楼	复交浙大人大大的硕士有把握嘛？感觉这就业行情，文科+本科确实不好找工作	-4.02358
4	3楼	现在外保比本校还困难（某种意义上）	1.577334
5	4楼	绩点如果比较靠前的话建议尝试外保，文科的话金融和法律就业会好一些，可以看看复交人的非法本法硕和商科项目，有的应该相对简单些，比如人大、复旦非法本法硕	3.563405
6	5楼	找找国外的全职predoc或者ra呢？phd大部分都是是有工资的，届时选择会大很多	8.229257
7	6楼	人文学科也有这些玩意吗？	1.56089
8	7楼	建议还是考公，当前的经济形势短期看还是比较严峻，你最好考回老家省会城市或者选调回去，省会公务员无论如何买房立足问题不大，北大文凭加上省会公务员婚姻问题不大	7.514857
9	8楼	香港一年还行的。身边有人香港一年硕找到了比较好工作的（记住千万不要考虑读博）	7.704368
10	9楼	别来4	-4.81025
11	10楼	华为要喝酒的	-1.98837
12	11楼	关键还是想好自己想要啥，不要从众，如果对薪资要求比较高，那就华为或互联网，华为客户经理如果发展得好确实薪资非常可观，而且也不会像互联网那么不稳定，而且也不会像互联网那么不稳定，而且也不会像互联网那么不稳定	-9.17673
13	12楼	从楼主的个人描述来看，如果真的挺喜欢社交的话我感觉当客户经理挺适合你的，关键还是你要对工作强度有心理预期，华为内部各个部门、不同的团队工作强度千差万别	-4.13015
14	13楼	辅修了经管，复交的经营够不上门槛，文科或许可以勉强一试	3.298535
15	14楼	谢谢建议，目前也是打算多手抓，因为确实不敢all in某一条赛道，目前打算先找一段日常实习+考一个标化英语成绩（已有英语成绩不太够用），保研的话要看这学期的成绩	2.169449
16	15楼	辅修了经管，但是没数学，经管方面的这种岗位太少了太卷了 人文社科的还没了解，再去看看，谢谢建议！	0.2921
17	16楼	嗯嗯，家人什么的也希望我考公，主要是出于稳定和婚姻，但是家庭养老负担太大了，害怕公务员薪资负担不了，婚姻的话也是因为自己家庭负担太大了，不太抱有期望	-8.42376
18	17楼	谢谢建议，香港一年硕已经了解过了，学费+生活费各种30-40w也需要，这笔钱已经很难拿出来了	4.535287
19	18楼	嗯嗯有了了解到，所以在观望	16.58718
20	19楼	请问学长/学姐可以细说吗（如果方便说的话）	-3.28746
21	20楼	互联网因为目前实习比较薄弱，暑期全被拒了，在找日常但目前也没有什么优势，这才想到华为，但是网上负面评价太多了，所以犹豫不决。目前打算去华为和进体制内	0.043298

北大未名

搜索版面、帖子、用户

帮助

用户名

下次自动登录

登录注册

找回密码

首页

版面收藏夹

版面目录

精华区收藏夹

首页 >> 版面目录 >> 八区: 三角地 >> 找工作啦(Job) >> 女生做华为客户经理还是进体制？

返回本版

1

2

3

4

下一页 >

1

/ 4

跳转

女生做华为客户经理还是进体制？

[复制链接]

分享:

楼主

Woxiangfacai [在线]

麦当脆薯饼

0.9 新手上路

发帖数: 37 原创分: 0

关注

1楼

<ASCIIArt>

人文社科大三，保研悬出国难，只能工作的话能做什么准备呢？
绩点绝对值还行（3.7+）但排名中游，只能拿到推免保研无缘top2，需要降级比较严重，且感觉现在来不及准备保研了；出国负担较为沉重，最多就是港一年或者欧陆无学费的，其他的真不行了，最多只有30w预算，这已经掏空家底了。因为家境实在窘迫（从出国也能看出来），加上感觉读人文社科的作用也不大，目前想着直接就业，有下面几个打算，还请大家提一些建议：
1. 考公进体制。党员+有一些学工奖励，还算有优势，但本科的话估计只能考一些非热门岗位，不过也能接受，个人物欲不是很强，但因为独生女需要负担父母养老可能比较辛苦。
2. 考教资去当老师。虽然是本科生，但是所在家乡当老师门槛很低，周边城市也是，薪资大概比1略少，但生活很舒服，wlb没问题，还离家近，可以照顾父母。这个主要是难度较小，但也面临舆论指责（XX毕业回来当老师这样子），且收入波动很大。
3. 卷互联网。其实已经试过了，因为之前只有一段实习还不是互联网的，暑期实习被好多互联网全挂了，到现在还没拿到offer，如果现在再卷一下找一段日常实习，带上之前的那段去秋招，继续找互联网，可能薪资会比体制内高一些，但也不会特别高（非产品非战略岗，市场职能类）
4. 其他方向就业。因为性格比较外向，不排斥社交甚至很喜欢，有考虑去做客户经理，比如华为这种，但了解到女生可能不太利好？比如酒桌文化等，以及奋斗者这种。但薪资确实客观，所以还在观望。

3. 数据结构与算法

3.1 数据选定

本组选定的是未名 BBS 中的评论，希望可以参考情绪评分方法针对未名 BBS 中的内容进行情绪评分。

3.2 数据处理

本组数据处理部分：首先由 gettop.py 爬取前十条的 url 和标题并保存；然后用 bbsWash.py 对最近未名 BBS 的评论进行清洗，得到以网页的 bid_threadid 命名的.csv 文件，储存在 csv 文件夹中。接着参考前述情绪评分方法，在 bbsNLP.py 中使用 SnowNLP 对清洗处理过的文本进行情绪评分，并保存原文件中。在 GUI 图形界面中，提供了查看统计、查看详情、查看网站三个接口。

3.3 评分过程

在进行此项目评分时，需要先下载以下缺省源：

```
pip install matplotlib
pip install webbrowser
pip install pandas
pip install openpyxl
pip install requests
pip install snownlp
pip install seaborn
```

在爬取评论进行评分时，运行 main.py，爬取热门前十的 bbs 版面，用以分析大学生心理健康与各种因素的关系分析并进行评分。

本小组希望可以尝试运用此模型来训练其他模型，来进行相关项目的拓展。本组认为此研究项目虽没有过多的直接用途，但是可扩展性强，可以与其他人文社会学科或是语言模型等进行交叉，提供情绪分析和观察的功能。

数据评分上面本组也做些调整，比起 SnowNLP 的直接评分，本组进行了简单的评分模型优化，即折合分 = 原始分 * 100 - 33，希望可以透过这样的改动来使情绪评分更直观。

4. 编码实现

4.1 数据获取

我们的想法一开始是爬取并分析热搜评论以及 Youtube 直播中的实时互动数据。但通过 api 爬取动态网页需要网页授权等很多网络相关技术，经过尝试并不能很好的解决。因此我们决定爬取静态网页的评论数据——未名 BBS。

第一版代码是根据本课程学到的用 requests 进行的网页源代码获取和用 re 进行的数据清洗进行网页爬取并得到具体的评论内容，并用文件操作保存到本地 csv 文件中。

4.2 添加 NLP 处理

接着，我们考虑处理这批数据。我们调查了几个基于中文分词进行情感分析的自然语言处理数据库，最终采取了开源的 SnowNLP 对我们爬取到的数据进行情感分析和情绪评估，并对每一条评论进行打分。

SnowNLP 的原理是对语境中的汉语词汇进行预打分，在处理语段时，对其进行汉语分词后取每个分词的平均分作为语段的情绪评分。很明显，这种简单的处理方式只能对简单而直接表达情感的语句进行处理，而对于反讽、欲扬先抑等复杂的情感表达方式并不能很好的识别。所以，这就适合讨论较为实际的 BBS 情感分析，而不适合微博、B 站、Youtube 等论坛里的评论区（大多阴阳怪气）进行情绪处理。

4.3 批次数据处理

如果每一个网页都需要用户提供网址作为输入数据，未免太过繁琐。于是我们添加了一个一次性批量爬取 10 个热度最高的帖子中的评论进行情绪评分，既方便直接查询又方便横向比较。

4.4 GUI 图形界面

为方便用户使用，我们添加了 GUI 图形界面。使用的是 python 中的 tkinter 库。为了美观，我们的 GUI 做出以下优化：

1. 增加了背景颜色和文本颜色
2. 使用全屏自适应
3. 优化了字体，汉字使用楷体书写，并选用合适的字体大小
4. 美化功能按键
5. 绑定窗口关闭时自动结束程序，从而防止重复运行报错。
6. 对柱状图进行了主题优化

5. 测试部署

5.1 开发工具和环境

开发语言使用 python，操作系统使用 Windows10，编译器使用 Python3.12。

经测试，本项目可以在 Windows10，Windows11，macOS，UbuntuLinux 系统中正常使用。

5.2 单元单项测试

每个文件中的 main 函数都是对该程序功能的测试，如果存在 bug，可以逐个文件调试，后期维护比较方便。

5.3 后期维护

本项目将持续维护和更新。如果您发现本项目存在漏洞，或者相位本项目贡献一个新的功能或进行改进，十分欢迎在 Github 上进行 PullRequest！

6. 软件使用说明

运行 main.py 文件即可显示主界面，如前图所示

运行时将自动在当前工作区新建两个文件夹 csv 和 ima，用于存放清洗文件和柱状图文件，

当关闭窗口时自动删除。

由于代码基于网页源码清洗来判断发帖内容，所以理论上可能会有些帖子的内容会引起报错或乱码。但这种情况只在恶意卡 bug（如发帖评论内容包含 html 格式的代码）的时候会出现，所以基本可以忽略。

7. 评价与总结

本研究通过对未名 BBS 评论区的情绪评估，揭示了大学生在不同时间段内的情绪状态及其变化规律。研究结果为学校在高压力时期更有效地分配心理疏导资源提供了数据支持，并为未来情绪分析模型的扩展应用奠定了基础。通过不断优化和改进情绪分析方法，可以更好地帮助学生应对心理压力，促进其心理健康发展。

另外一方面，巩固了本课程所学的网络爬虫与信息处理、文件操作、GUI 图形界面设计等知识。并且比较成功的将课堂所学应用到实际项目中，激发了我们小组同学的学习兴趣，项目制作过程中也有很多新的问题，解决这些问题使我们对知识的使用更加灵活。

8. 小组成员信息与项目分工

序号	姓名	学号	email	完成的任务
1	吴宇航	2310301128	2310301128@stu.pku.edu.cn	选题讨论，整体构思， 报告撰写，写代码
2	金旻河	2100093023	2100093023@stu.pku.edu.cn	选题讨论，整体构思
3	陳妍安	2200016609	2200016609@stu.pku.edu.cn	选题讨论，整体构思， 报告撰写
4	果果	2000093014	2000093014@stu.pku.edu.cn	选题讨论，整体构思， 报告撰写

9. 参考文献

[1] 钟雪灵等.Python 金融数据挖掘[M].北京：高等教育出版社，2020：132.

[2] 基于汉语分词的文本情感分析系统 [SnowNLP](#)