

**新疆大学课程论文（设计）、学年论文评分表**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 题 目 | B站科学科普视频的弹幕情感分析 | | | | |
| 作 者 | 侯晓薇 | 专业年级 | 网安20-3 | 指导教师 | 程述立 |
| 指导教师评语及  评分建议 | 该生积极完成网络舆情分析课程作业并参与课程研讨，上课认真听讲，论文格式较为规范，图表清晰，文字表述通顺，用语符合专业要求，逻辑性较强，课程设计较为合理，论述过程较为详细。整体而言，论文结构科学合理，逻辑思路清晰，观点表达准确，语言流畅，论述方法清晰合理，参考文献与课程主题符合，论文内容与网络舆情分析课程内容相符。  根据网络舆情分析的课程要求，结合该生考勤和作业情况，平时成绩为： （ ）分，根据课程论文撰写质量和课程论文工作量，课程论文成绩评定为：（ ）分，综合成绩为平时成绩占30%，课程论文占70%，该生综合成绩评定为： （ ）分。  根据以上意见，建议该生课程论文最终成绩等级为：（ ）。  指导教师：  2023 年 6 月 18 日 | | | | |
| 院  （部）  或  教  研  室  意  见 | 同意指导老师意见。  学院或教研室主任：  年 月 日 | | | | |

摘 要

B站创建之初的定位是围绕动画、漫画、游戏等内容进行原创与分享的网站，后逐渐进入大众视线，演变为内容丰富，创作氛围良好的年轻潮文化娱乐社区。现在的B站不仅受到喜爱动漫、游戏的年轻人的欢迎，又追随大众喜好加入了如科技、音乐、鬼畜、舞蹈等专区，而其最具特色的便是其弹幕系统，形形色色的弹幕在观看之余为用户探讨动画情节,科普相关知识等提供了便利。而弹幕本身也包含了发送者观看视频时的即时感情和观点，当up主制作原创视频上传后也可以通过弹幕了解到用户对该内容的观点和情感走向，以此也为up主制作更多优质视频、平台规范视频内容提供了依据。

弹幕文本相较于微博等评论性质的文本不同,不仅包含大量的网络流行语和颜文字表情，还存在大量“同词不同义”的情况，这些问题都给弹幕文本情感分析工作带来了挑战。但目前国内做弹幕情感分析的文章相对较少，且大多数文章倾向于学术研究，数据可视化程度低，应用性不强，再者弹幕本身具有简洁性、口头用语、网络词汇较多，且用语较于评论文本更加不规范等问题，这些都使得进行准确的弹幕情感分析变得较为不易。基于以上问题，本文主要研究内容和工作如下: 首先是通过爬虫技术对B站视频的弹幕爬取工作，对于爬取到的数据进行处理，处理后的数据利用WordCloud绘制词云图，再利用自然语言处理的SnowNLP绘制情感分析图。

**关键词：**B站;爬虫；WordCloud; NLP

**Abstract**

At the beginning of its creation, Bilibili was positioned as a website for original creation and sharing of content such as animations, comics, and games. Later, it gradually entered the public eye and evolved into a rich and creative young cultural and entertainment community. Nowadays, Bilibili is not only popular among young people who love anime and games, but also follows the public's preferences by joining special zones such as technology, music, ghost animals, and dance. Its most distinctive feature is its barrage system, which provides convenience for users to explore animation themes and popularize related knowledge while watching. And the barrage itself also contains the real-time emotions and opinions of the sender when watching the video. After uploading the original video created by the up host, the user's views and emotional trends on the content can be understood through the barrage, which also provides a basis for the up host to produce more high-quality videos and standardize the video content of the platform.

The bullet screen text is different from the comment text such as microblog. It not only contains a large number of network catchwords and Emoticon 'expressions, but also has a large number of situations of "the same word with different meanings". These problems have brought challenges to the emotion analysis of bullet screen text. However, at present, there are relatively few articles on bullet screen emotion analysis in China, and most of the articles tend to academic research. The degree of Data and information visualization is low, and the applicability is not strong. Moreover, bullet screen itself has many problems such as simplicity, oral language, network vocabulary, and the language is more non-standard than the review text. All these make it difficult to conduct accurate bullet screen emotion analysis. Based on the above problems, the main research contents and work of this paper are as follows: First, the bullet screen crawling work of station B video is carried out through crawler technology, and the crawled data is processed. The processed data is used to draw a word cloud map using WordCloud, and then the emotion analysis map is drawn using SnowNLP of Natural language processing.

**Key Words:** Bilibili; Reptiles; WordCloud; NLP

**目 录**

[第1章 绪论 1](#_Toc137826793)

[1.1 研究背景及意义 1](#_Toc137826794)

[1.2 国内外研究现状 2](#_Toc137826795)

[1.4 本文研究内容及章节安排 3](#_Toc137826796)

[1.4.1 论文研究内容 3](#_Toc137826797)

[1.4.2 论文章节安排 3](#_Toc137826798)

[第2章 相关理论和技术介绍 5](#_Toc137826799)

[2.1 爬虫相关理论与技术 5](#_Toc137826800)

[2.2文本情感分析相关理论与技术 7](#_Toc137826801)

[2.2.1 SnowNLP情感分析技术 8](#_Toc137826802)

[2.2.2jieba分词 10](#_Toc137826803)

[2.3本章小结 10](#_Toc137826804)

[第3章 B站科学科普视频弹幕的情感分析 11](#_Toc137826805)

[3.1 实验环境 11](#_Toc137826806)

[3.2 实验过程设计 12](#_Toc137826807)

[3.3 实验结果及分析 12](#_Toc137826808)

[3.3.1 数据收集与预处理 12](#_Toc137826809)

[3.3.2 词云分析 16](#_Toc137826810)

[3.3.3 情感分析 18](#_Toc137826811)

[3.4 本章小结 19](#_Toc137826812)

[第4章 总结和展望 20](#_Toc137826813)

[4.1 工作总结 20](#_Toc137826814)

[4.2 工作展望 20](#_Toc137826815)

[参考文献 21](#_Toc137826816)

[致谢 22](#_Toc137826817)

第1章 绪论

1.1 研究背景及意义

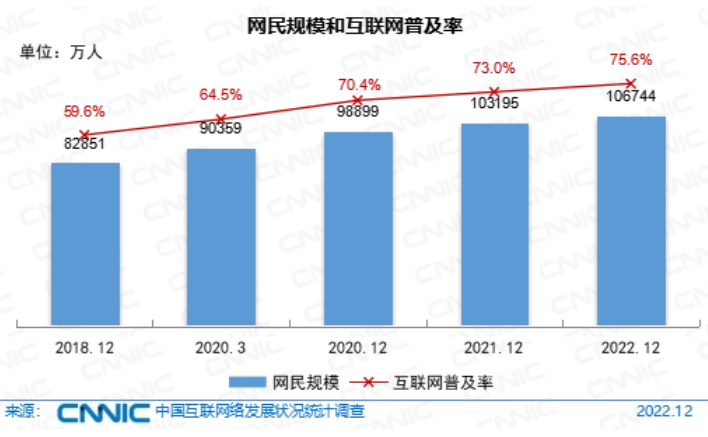
2023年3月2日，中国互联网络信息中心（CNNIC）发布第51次《中国互联网络发展状况统计报告》（以下简称《报告》）。《报告》显示，截至2022年12月，我国网民规模达10.67亿，较2021年12月增长3549万，互联网普及率达75.6%。如图1-1所示。

图1-1网民规模与互联网普及率

B站虽然在在线视频市场中非占据主流地位，但是凭借其主推动漫、游戏视频等优势，也在激烈的市场竞争中占据了一席之地。在B站最具特色的是引领了弹幕这种独特的交流方式，平均每分钟有数万人发送弹幕，日均弹幕发送数量也达到百万级。可见，用户在观看视频之余发送弹幕的行为越来越常态化。与此同时，对弹幕的情感分析技术具有较好的市场应用价值。

国内在线视频市场的快速发展，众多网民可通过视频观看时事新闻、刷剧追剧并通过弹幕参与各种事件的讨论，从个人分享到商业活动，再到全球性的重大事件,B站俨然成为大众情感表达的社交平台。通过对B站中用户发布的弹幕进行情感分析对于个人、up主以及平台的发展都具有深远的意义。

与其他视频网站相比，B站的视频用户互动率更高，热度较高的作品出现弹幕刷屏的现象十分普遍，up主可以通过流量变现获取一定的收益。与此同时，数十个不同的兴趣分区保证优质的内容创造者更容易受到大众的追捧，这在某种程度上启发了原创视频作者的创作活力，也有利于大众创新。并且用户在该视频上发布弹幕的褒贬在某种程度上了说明了该视频制作的优劣，这也就为up主创作更优质的视频创作或进行创作转型提供了方向。

最后,平台了解弹幕的情感倾向也为平台取缔低俗视频提供了依据。当某一视频中的含有负面情绪的弹幕较多时，平台可以对该视频重新进行审核，若不符合平台相关规范,或是传播了网络负能量或是含有引战言论，就应及时对该视频下架处理，甚至对相关视频上传者做出处罚。平台同时也可以根据弹幕判断用户喜好，不断拓展社区内容的深度与广度，从而更多吸引不同圈层的受众。而这些构建了平台健康运营的环境,提供了和谐创作的氛围，同时也吸引了更多优质原创内容的作者入驻B站。

1.2 国内外研究现状

目前国内外对弹幕文本情感分析所做的工作主要有基于情感词典、弹幕常用词词典、注意力机制的神经网络、双通道卷积神经网络、奇异值分解算法的卷积神经网络五个方面[2]。

基于情感词典的情感分类方法需要首先人工创建情感词典,囊括大量情感词汇,再从弹幕中提取情感词汇并与情感词典中的基准词相匹配，最终得出整条弹幕的情感倾向。郑飚飚⑶等人通过建立情感词典来进行弹幕情感分析，并对评论文本进行情感词抽取及情感值计算，并结合时间序列进行分析。赵天锐[4]等提出了一种将韩国语通用情感词典拓展为影评领域情感词典的方法,以网络电影评论为基础构建语料库，通过合并通用情感词典和已判别情感倾向的情感词,得到领域情感词典。Jang 等[5]基于word2vec方法构建了特定主题的情感词典，具体做法是先选定种子词，然后用word2vec方法将种子词转化为向量形式并通过余弦相似度选出候选情感词。

弹幕常用词词典与情感词典相似,都需要手动构建情感词典，但洪庆和王思尧等[6]在使用弹幕常用词词典的基础上通过改进传统的k-means聚类算法，对弹幕常用词词典与情感词典相似,都需要手动构建情感词典，,以此帮助我们了解观看特定类型视频的观众在情感上的异同点。

长短期记忆网络和卷积神经网络都是使用深度学习的方法进行弹幕情感分析。Kalchbrennerf[7]等提出了动态CNN模型，该模型采用动态K-Max 池化，区别于传统的最大池化只保留一个重要信息。庄须强，刘方爱等[8]基于注意力机制挖掘出整个弹幕评论中的情感关键词;然后利用LSTM模型有效结合视频中前后弹幕评论的情感依赖关系，最终提取出基于主题的“高光”视频。Ll ping 等1DJ使用双通道卷积神经网络对文本进行情感分析，其中一个通道为字向量，另一个通道为词向量,解决了单通道卷积神经网络视角单一以及不能充分学习到文本特征信息的问题。

从以上发展过程不难看出，文本情感分析研究正在不断深入，人的情感本身就具有复杂性，这也侧面说明了文本情感分析工作还有很长的路要走。

1.4 本文研究内容及章节安排

为了能满足相应用户的使用需求,通过对相关用户调查了解并在网络上查询了大量与情感分析有关的功能需求，以用户需求为出发点.本文的实验结果的面向的对象主要是B站的用户和up主。对于用户来说，用户更倾向会选择更加优质的视频来观看，这导致他们很难在海量视频中抉择;而对于up主来说，他们会关注自己上传原创视频的质量和观看用户对此视频的褒贬，由于视频中出现的海量弹幕和评论，他们很难从中提取有用信息来判断大众对该原创视频的评价，同时他们在原创过程中也经常对当前热点不甚了解，导致创作无从下手。若当创作者能够较为准确的得到观众的情感状态，对于创作是百利而无一害的。

本文研究的内容则是通过使用python的爬虫技术，收集到观众在观看视频的过程中发送的实时弹幕，顶部弹幕和滚动弹幕的收集。对收集到的数据进行清洗和处理等得到数据集，将数据可视化展现出来。通过使用一些技术绘制出弹幕的词云图和情感分析图。创作者可以通过可视化的结果来进行创作上的调整做出高质量的视频等

1.4.1 论文研究内容

在本文中，第1章为研究背景意义及国内外研究现状，第2章是相关理论支撑和技术指导，第3章则是具体实验过程具体研究内容描述如下：数据收集，词云分析，情感分析，第4章是总结和展望。

1.4.2 论文章节安排

本文章节安排如下：

第1章：绪论。首先探讨研究背景及意义，然后阐述国内外研究现状，接下来叙述课题来源，最后总结研究内容并安排论文章节。

第2章：相关的理论和技术。介绍爬虫相关的理论和技术，介绍文本情感分析的技术，介绍了本文在研究过程中使用到的两种技术。

第3章：B站科学科普视频弹幕的情感分析。首先介绍了实验环境，实验设计，及实验结果，词云分析结果，和情感分析的结果。

第4章：总结与展望。首先总结全文研究工作，然后展望下一步研究计划。

第2章 相关理论和技术介绍

前面章节主要阐述了B站弹幕情感分析的研究现状，其中包括基于情感词典、弹幕常用词词典、注意力机制的神经网络、双通道卷积神经网络、奇异值分解算法的卷积神经网络五个方面的研究现状。区别于前面章节内容，本章主要阐述与本文工作相关的理论基础和技术。研究的主要内容是先利用爬虫技术获取道视频弹幕的数据信息，对信息进行处理，最后弹幕进行情感分析，其目标是充分利用弹幕数据进一步分析观众的情感状态。因此，本章内容在表述顺序上，先阐述了关于爬虫的相关理论与技术，接着开始介绍文本分析的相关理论与技术。

2.1 Python爬虫相关理论与技术

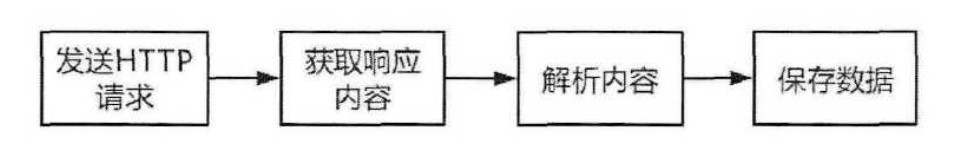
爬虫通俗来讲就是通过程序模拟预览器请求站点行为，把相关数据抓取下来,然后将自己所需的数据提取并存放起来。Python爬虫首先向目标站点发起请求，如果得到服务器的响应，则通过正则表达式或第三方解析库解析数据内容，最后将数据保存至数据库中，具体流程如图2-1所示。

图2-1爬虫基本流程

Python爬虫拥有多种数据获取方法，并各自具有其优缺点，分别是同步、异步、并发和 Scrapy框架。同步方法思路简单，容易实现，但效率不高爬取数据较慢;并发的爬虫爬取速度更快，但抓取开销也更大，会消耗更多的时间成本;异步方法可以获得更高的抓取效率，但需掌握异步编程，需要一段时间的学习,并不适合初学者使用;使用Scrapy框架在爬取完成后可以自动生成.csv文件，且支持异步，并发，容错性好，但如果需要频繁地修改中间件，则Scrapy框架就显得并不灵活，而且它在速度上也没有超过异步爬虫。于此，爬虫提供了多种方法,开发者可以依据实际情况选择适用自己的方法,给开发者提供了多元化的选择。

Python 爬虫架构主要由五个部分组成，分别是调度器、URL管理器、网页下载器、网页解析器、应用程序（爬取的有价值数据）。

调度器：相当于一台电脑的CPU，主要负责调度URL管理器、下载器、解析器之间的协调工作。

URL管理器：包括待爬取的URL地址和已爬取的URL地址，防止重复抓取URL和循环抓取URL，实现URL管理器主要用三种方式，通过内存、数据库、缓存数据库来实现。

网页下载器：通过传入一个URL地址来下载网页，将网页转换成一个字符串，网页下载器有urllib2（Python官方基础模块）包括需要登录、代理、和cookie，requests(第三方包)

网页解析器：将一个网页字符串进行解析，可以按照我们的要求来提取出我们有用的信息，也可以根据DOM树的解析方式来解析。网页解析器有正则表达式（直观，将网页转成字符串通过模糊匹配的方式来提取有价值的信息，当文档比较复杂的时候，该方法提取数据的时候就会非常的困难）、html.parser（Python自带的）、beautifulsoup（第三方插件，可以使用Python自带的html.parser进行解析，也可以使用lxml进行解析，相对于其他几种来说要强大一些）、lxml（第三方插件，可以解析 xml 和 HTML），html.parser 和 beautifulsoup 以及 lxml 都是以 DOM 树的方式进行解析的。

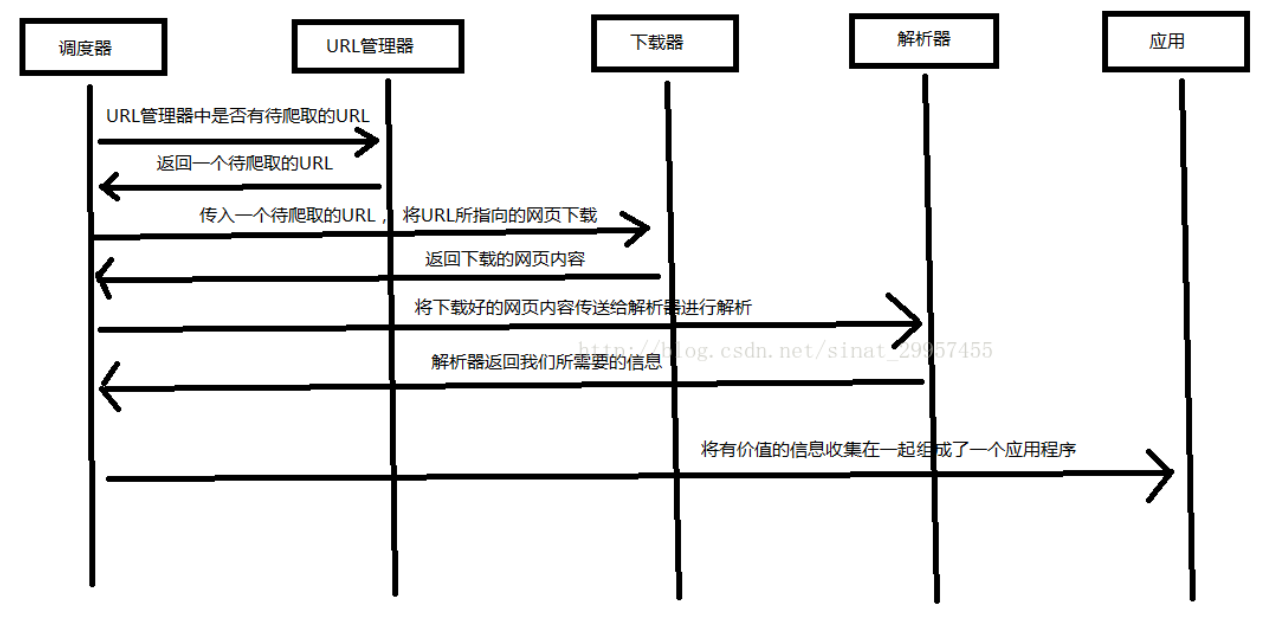
应用程序：就是从网页中提取的有用数据组成的一个应用。框架如图2-1所示。



图2-2 python爬虫框架

2.2文本情感分析相关理论与技术

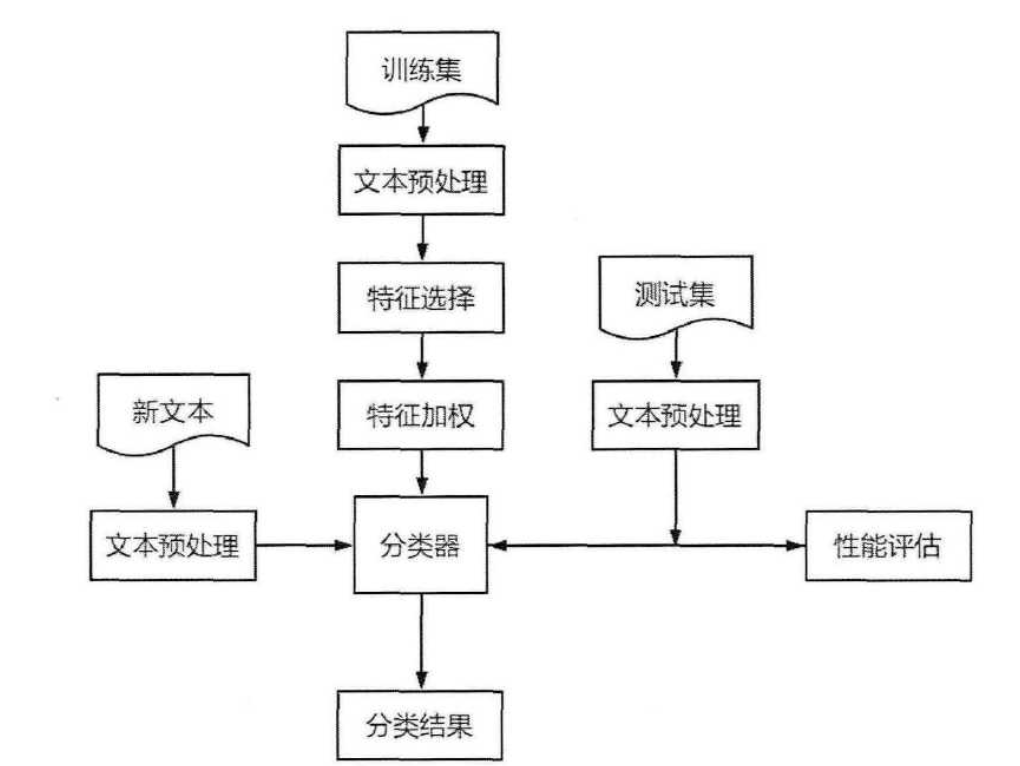
文本情感分类过程就是把收集到的带有情感色彩的文本进行分析、处理等过程。如图2-2所示，文本情感分类过程大致分为五步:第一步进行文本预处理，将文本主要内容提取出来，去除与情感分类无关的词语，主要包括分词、去除停用和过滤非法字符等;第二部就是通过特位挺代，土从为里的分类特征,第三部了一个高维特征空间，再从中选择一小部分特征作为分类器的分类特征;第三部是进行特征加权，根据特征文本的表达程度不同，赋予其相对应的权重，这样可以有效抑制噪声，常用的特征加权方法有布尔权重，词频权重，TFIDF 权重;最后一步则是为文本分类选择适合的分类器，常见的有线性回归、SVM、贝叶斯分类器、决策树和K近邻分类器等。

图2-2文本分类

2.2.1 SnowNLP情感分析技术

自然语言处理，即实现人机间自然语言通信，或实现自然语言理解和自然语言生成是十分困难的。造成困难的根本原因是自然语言文本和对话的各个层次上广泛存在的各种各样的歧义性或多义性(ambiguity)。

自然语言处理是计算机科学领域与人工智能领域中的一个重要方向。它研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法。自然语言处理是一门融语言学、计算机科学、数学于一体的科学。因此，这一领域的研究将涉及自然语言，即人们日常使用的语言,所以它与语言学的研究有着密切的联系，但又有重要的区别。自然语言处理并不是一般地研究自然语言，而在于研制能有效地实现自然语言通信的计算机系统，特别是其中的软件系统。因而它是计算机科学的一部分。自然语言处理(NLP)是计算机科学，人工智能，语言学关注计算机和人类(自然)语言之间的相互作用的领域。

自然语言处理(NLP)是计算机科学，人工智能，语言学关注计算机和人类(自然)语言之间的相互作用的领域。因此，自然语言处理是与人机交互的领域有关的。在自然语言处理面临很多挑战，包括自然语言理解，因此，自然语言处理涉及人机交互的面积。在NLP诸多挑战涉及自然语言理解，即计算机源于人为或自然语言输入的意思，和其他涉及到自然语言生成。

现代NLP算法是基于机器学习，特别是统计机器学习。机器学习范式是不同于一般之前的尝试语言处理。语言处理任务的实现，通常涉及直接用手的大套规则编码。

许多不同类的机器学习算法已应用于自然语言处理任务。这些算法的输入是一大组从输入数据生成的"特征"。一些最早使用的算法，如决策树，产生硬的if-then规则类似于手写的规则，是再普通的系统体系。然而，越来越多的研究集中于统计模型，这使得基于附加实数值的权重，每个输入要素柔软，概率的决策。此类模型具有能够表达许多不同的可能的答案，而不是只有一个相对的确定性，产生更可靠的结果时，这种模型被包括作为较大系统的一个组成部分的优点。

自然语言处理研究逐渐从词汇语义成分的语义转移，进一步的，叙事的理解。然而人类水平的自然语言处理，是一个人工智能完全问题。它是相当于解决中央的人工智能问题使计算机和人一样聪明，或强大的AI。自然语言处理的未来一般也因此密切结合人工智能发展。

SnowNLP是一个功能强大的中文文本处理库，它囊括了中文分词、词性标注、情感分析、文本分类、关键字/摘要提取、TF/IDF、文本相似度等诸多功能，像隐马尔科夫模型、朴素贝叶斯、TextRank等算法均在这个库中有对应的应用。SnowNLP是一个python写的类库，可以方便的处理中文文本内容，是受到了TextBlob的启发而写的，由于现在大部分的自然语言处理库基本都是针对英文的，于是写了一个方便处理中文的类库，并且和TextBlob不同的是，这里没有用NLTK，所有的算法都是自己实现的，并且自带了一些训练好的字典。

Snownlp主要功能包括：

中文分词（算法是Character-Based Generative Model）

词性标注（原理是TnT、3-gram 隐马）

情感分析

文本分类（原理是朴素贝叶斯）

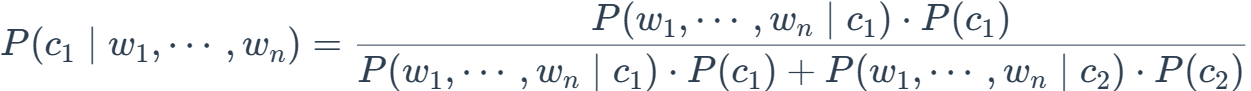
转换拼音、繁体转简体

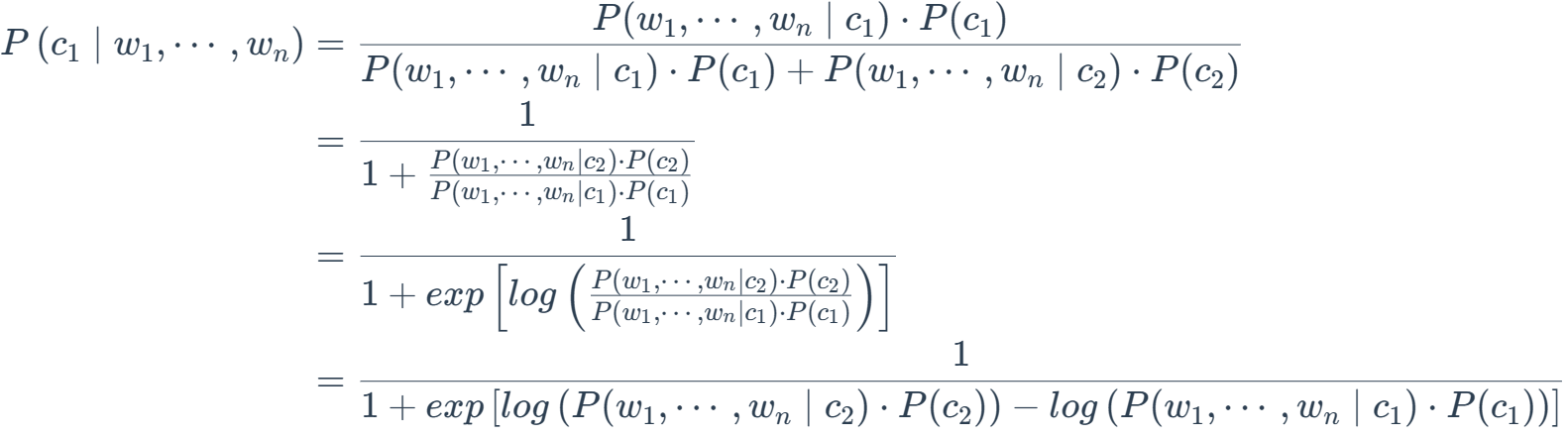
提取文本关键词（原理是TextRank）

提取摘要（原理是TextRank）、分割句子

文本相似（原理是BM25）

SnowNLP情感分析也是基于情感词典实现的，其简单的将文本分为两类，积极和消极，返回值为情绪的概率，越接近1为积极，接近0为消极。

预测过程运用到了以下公式：

对上述公式简化：

将分母化简为1时：



2.2.2jieba分词

jieba 是目前表现较为不错的 Python 中文分词组件，它主要有以下特性：中文文本需要通过分词获得单个的词语，jieba需要额外安装,jieba库提供三种分词模式。jieba库的分词原理：利用一个中文词库，确定汉字之间的关联概率，汉字间概率大的组成词组，形成分词结果。

jieba库支持四种分词模式：精确模式、全模式、搜索引擎模式、paddle模式，并且支持繁体分词，以及自定义词典。具体介绍：

精确模式：试图将句子最精确地切开，适合文本分析；

全模式：把句子中所有的可以成词的词语都扫描出来, 速度非常快，但是不能解决歧义；

搜索引擎模式：在精确模式的基础上，对长词再次切分，提高召回率，适合用于搜索引擎分词。

Jieba的算法如下：

一是基于前缀词典实现高效的词图扫描，生成句子中汉字所有可能成词情况所构成的有向无环图 (DAG)

二是采用了动态规划查找最大概率路径, 找出基于词频的最大切分组合

三是对于未登录词，采用了基于汉字成词能力的 HMM 模型，使用了 Viterbi 算法。

2.3本章小结

在对B站的视频弹幕情感分析研究过程中爬虫相关的技术和文本情感分类的技术和理论都为本文的研究提供了理论的支撑和技术的指导。在数据收集部分主要介绍了python爬虫的理论知识；在情感分析部分主要介绍了使用到的两种技术SnowNLP技术和jieba分词技术。

第3章 B站科学科普视频弹幕的情感分析

第2章主要讲述了本文在研究过程中使用到的一些技术具体描述如下：本文首先介绍有关Python爬虫技术，文本分类和SonwNLP技术的相关背景知识，然后在本章实现相关的技术得到科学科普视频弹幕的词云图，情感分析图以及柱状图。，接下来通过实验结果进行逐一展示，最后完成本章工作的小结。

3.1 实验环境

本文研究过程中的软件环境主要是Pycharm2021.3.1版本软件的Python3.7环境，并在Anaconda下安装Python依赖。如表3-1所示。

表3-1

|  |  |
| --- | --- |
| 依赖名称 | 版本号 |
| pandas | 2.0.3 |
| requests | 2.24.0 |
| lxml | 4.6.1 |
| jieba | 0.42.1 |
| matplotlib | 3.3.2 |
| snownlp | 0.12.3 |
| numpy | 1.8 |

3.2 实验过程设计

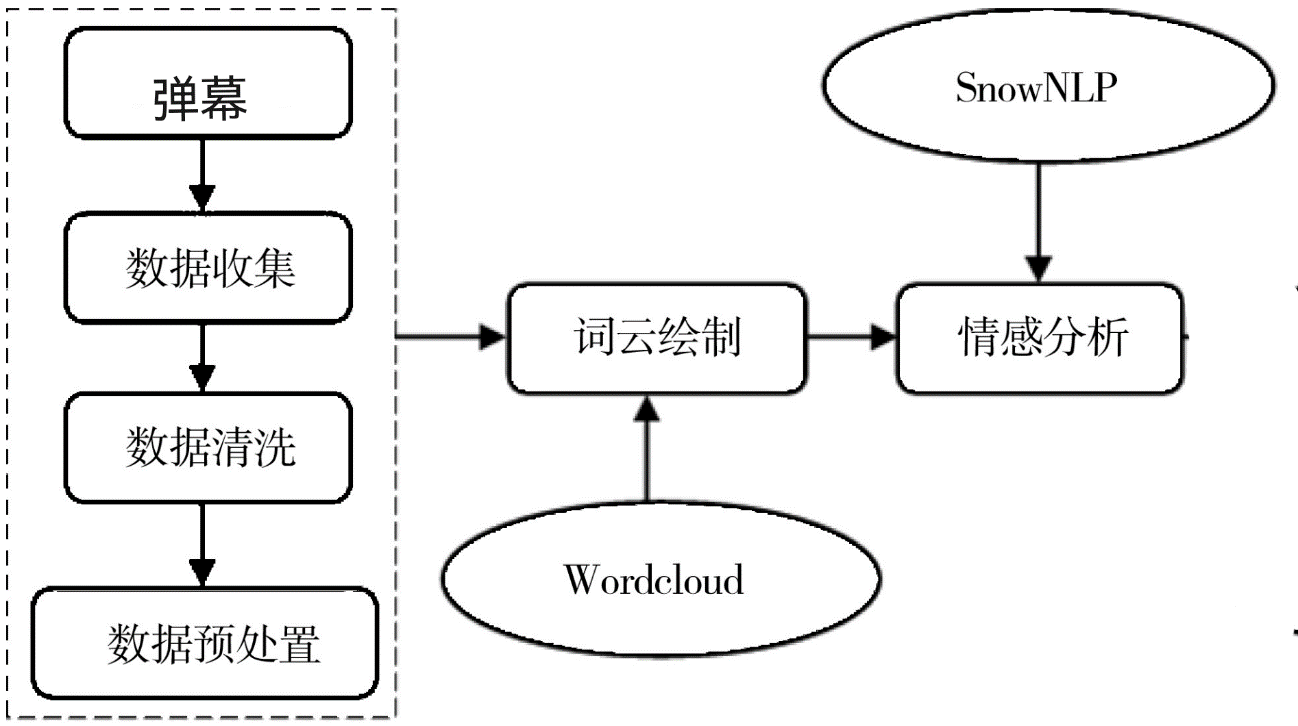
本文是对于B站科学科普视频的弹幕情感分析主要涉及到爬虫部分，词云图绘制和情感分析三部分，其中爬虫部分是由先弹幕爬取收集数据，再进行数据清洗和数据预处理。如图3- 2所示。

图3-2实验过程

3.3 实验结果及分析

对实验的爬虫部分和情感分析部分的可视化结果。

3.3.1 数据收集与预处理

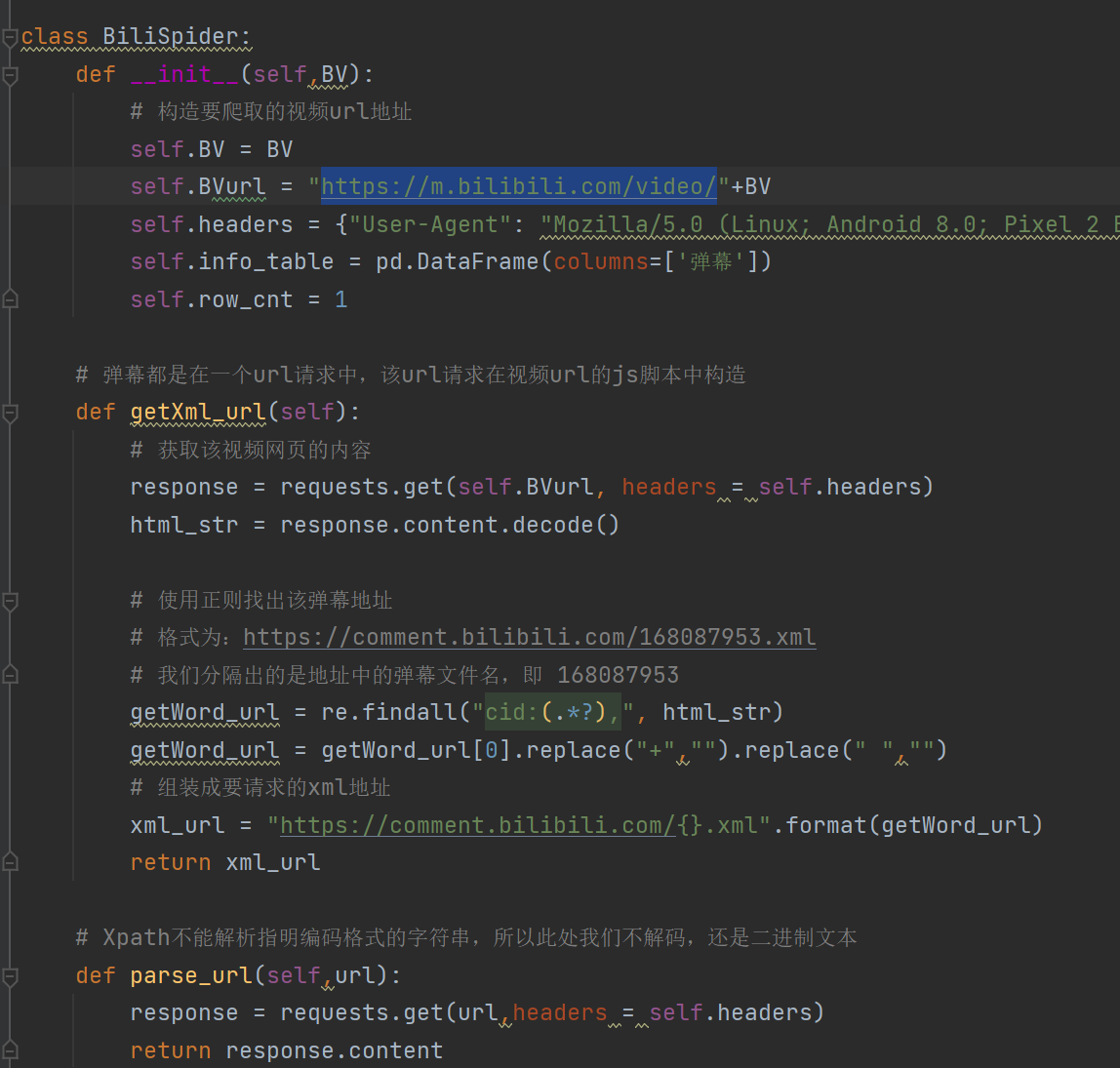
使用python爬取B站科学科普视频的弹幕的文本数据，获得分析所需的原始语料由爬虫爬取的评论数据，存在大量噪声数据，像常用表情符、键入特殊符、网页标签等，因此,文本数据需要清洗，通过jieba分词、 去除停用词、词性标注等方式,去除情感表达不明显或矛盾的内容。爬虫代码部分如图3-3-1所示。

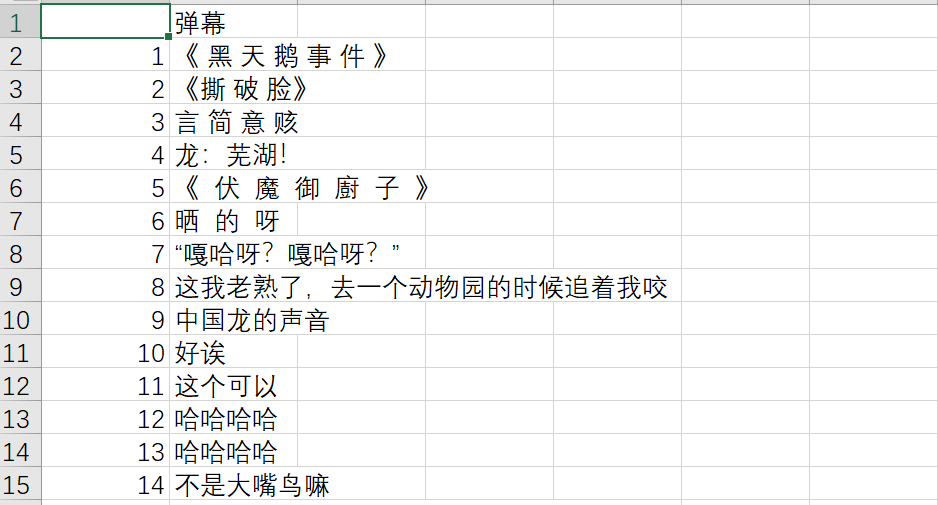
图3-3-1爬虫代码（一）



图3-3-2爬虫代码（二）

对于B站科学科普视频弹幕的爬取结果如图3-3-3所示。

图3-3-3科学科普视频的详细视频弹幕

以科学科普视频为例，弹幕的爬取结果如图3-3-4所示。

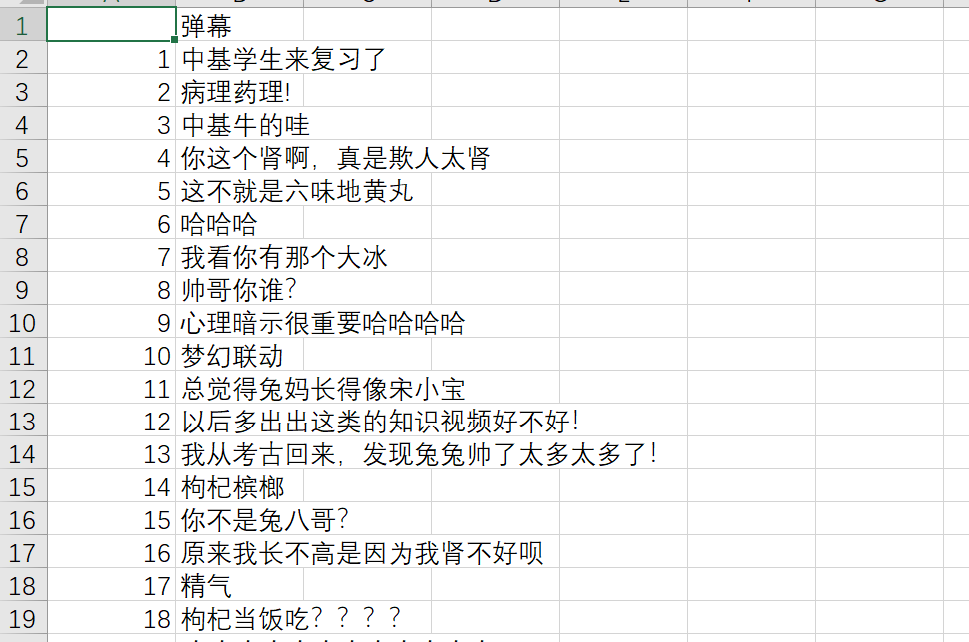
图3-3-4科学科普视频1结果图

图3-3-4科学科普视频2结果图

3.3.2 词云分析

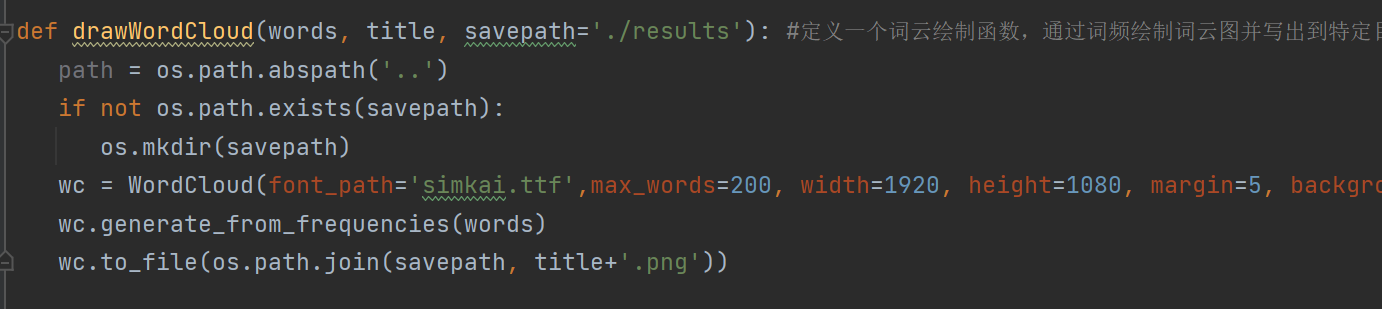
根据得到的弹幕运用TexBlob中的WordCloud子包对预处理后的数据进行词云绘制，出现次数较高的词会以可视化的方式呈现.借助词云图能高效地排除低频低质的信息，弱化非关键区域，突出呈现关键信息。"词云"就是通过形成"关键词云层"或"关键词渲染"，对网络文本中出现频率较高的"关键词"的视觉上的突出。词云图过滤掉大量的文本信息，使浏览网页者只要一眼扫过文本就可以领略文本的主旨。wordcloud库是python中的一个第三方库，在本文中词云图的绘制是直接调用了其库函数进行绘制的。代码如图3-3-5所示。

图3-3-5词云图代码部分

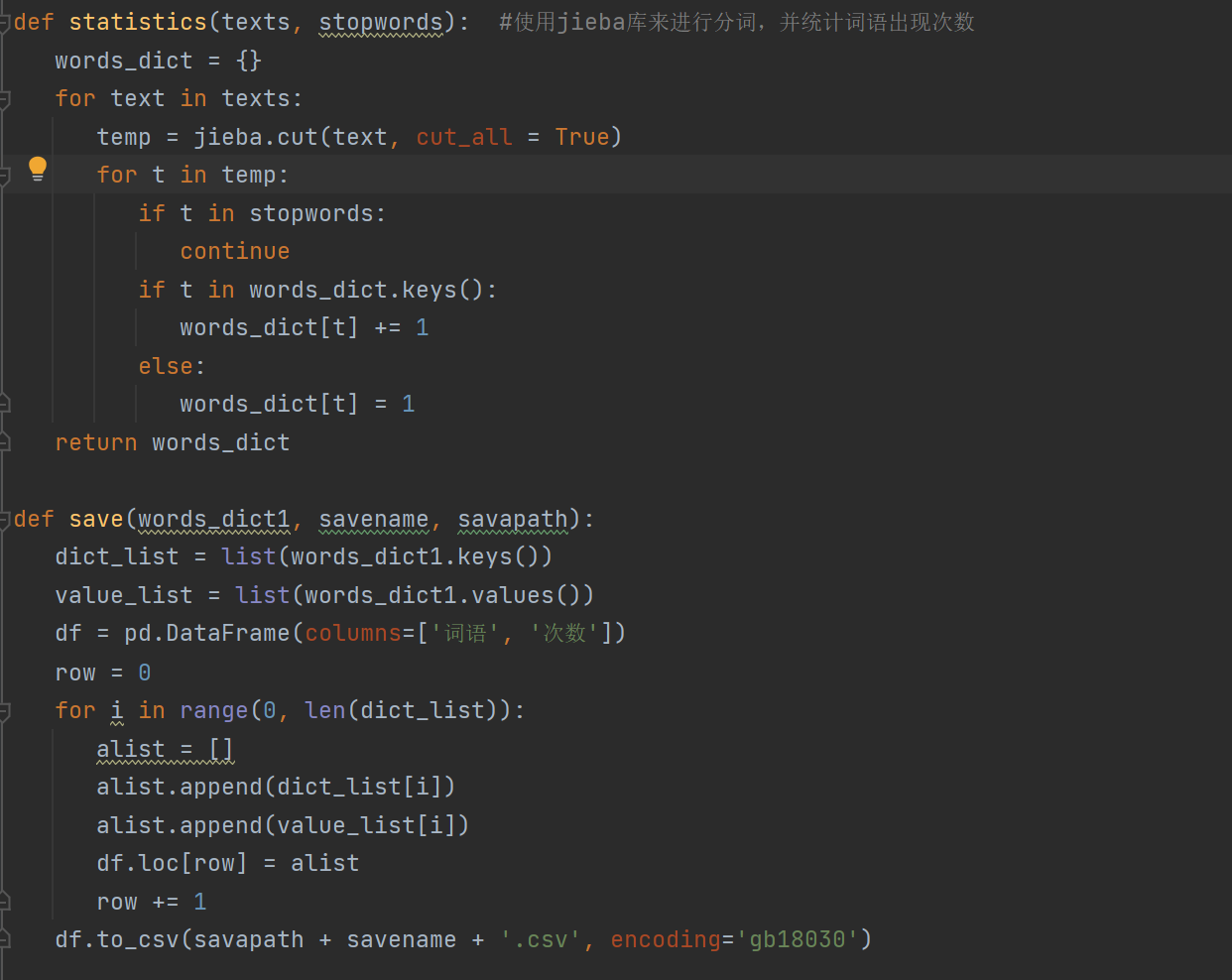
词频统计部分的代码如图3-3-6所示。

图3-3-6词频统计代码图

以科学科普视频为例，弹幕词云图如图3-3-7所示。

图3-3-7科学科普视频的词云图

3.3.3 情感分析

首先通过TextBlob中的SnowNLP子包来计算每条弹幕的情感得分，并以可视化的方式呈现。

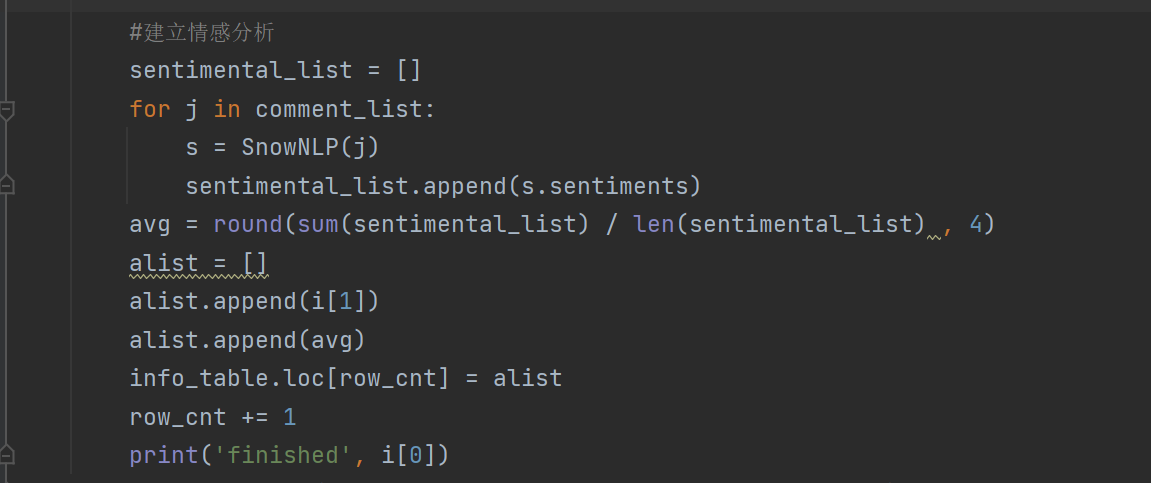
代码部分如图3-3-8所示。

图3-3-8情感分析图代码部分

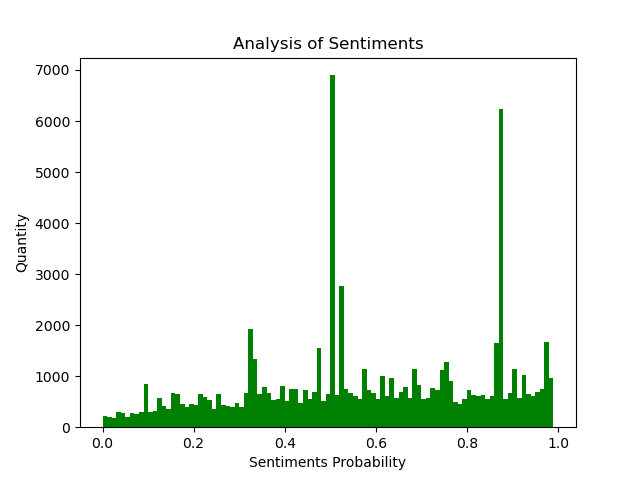
以科学科普视频弹幕为例得到的情感分析图如图3-3-8所示。

图3-3-8科学科普视频的情感分析图

3.4 本章小结

在本章内容中主要是操作过程与得到结果的展示。主要包含两个部分的实验和结果。在本章开头介绍了实验操作的软件环境以及依赖关系的版本号。在python爬虫的部分主要展示了代码部分。本文主要是针对科学科普视频的弹幕进行爬取，将得到的弹幕进行简化处理。得到数据后利用第2章提到的技术进行可视化展出，包括词云图和情感分析图。

第4章 总结和展望

4.1 工作总结

在互联网飞速发展的今天，B站视频走进了越来越多人的视野中。人们玩手机看视频的时间也越来越久，弹幕的文化至关重要，对于B站弹幕的情感分析也很重要，可以从中了解到观众的感受。这次我们就来聊一聊B站的弹幕文化。B站弹幕文化有着它独特的魅力，可以说是对B站文化的一种补充。现在许多人都会用B站来进行学习，这里也成为了学习知识的平台。在视频中，我们可以看到一些有趣的弹幕，可以与视频作者进行互动，了解到自己所不了解的知识。好的弹幕能够带给人积极正向的情绪，带给大家一些能量，当然消极的弹幕会给观众带来不好的观感同时对于创作者来说也是会造成不良影响。本文研究内容主要是对于弹幕的情感分析，主要针对科普视频，从弹幕的情感分析来看大部分人都是较为中立的状态，但是仍然有较为极端的情绪态度。

4.2 工作展望

本文对B站的科学科普视频弹幕中情感数据的特点及规律进行了分析，研究发现：科学科普视频弹幕的情感数据在整体上呈现出两极分化的特点；科学科普视频弹幕中情感分布呈现出明显的层次结构，这与弹幕文本在用户中传播时形成的层次结构有关；在科学科普视频弹幕中，用户会对不同类型的弹幕进行情感分类，但分类结果并不完全一致。当然本文的研究是基于随机挑选视频进行的情感分析，会存在很多不足之处，当然也可以通过弹幕能够了解到观众对于科学科普类视频的一些情感状态。对于创作者而言是一种优秀的工具。

未来，可从以下几个方面进一步研究B站的科学科普视频弹幕的情感数据：一是对B站的科学科普视频弹幕进行更加全面地采集；二是进一步完善B站的科学科普视频弹幕中情感数据的采集方法；三是基于B站的科学科普视频弹幕进行更加深入的情感分析。

参考文献

1. B站日活用户突破6500万成国内第三大长视频平台[J.中国有线电视，2021(08): 859.
2. 曾诚,温超东，孙瑜敏,潘列,何鹏.基于ALBERT-CRNN的弹幕文本情感分析[叮.郑州大学学报(理学版),2021,53(03): 1-8.
3. 郑随飚,徐健，肖卓.情感分析及可视化方法在网络视频弹幕数据分析中的应用[J.现代图书情报技术,2015(11): 82-90.
4. 赵天锐,刘晨阳.基于深度学习的韩国语影评情感词典构建[J.信息技术与信息化, 2021(01): 250-253.
5. Jang H S,K Y Jeong, E Y Jang.Efficient method to generate sentiment vocabularyfor specific topic based on Word2Vec [J]J. Proceedings of Korean Institute ofInformation Scientists and En-gineers, 2017: 652-654.
6. 洪庆，王思尧,赵钦佩，李江峰,饶卫雄.基于弹幕情感分析和聚类算法的视频用户群体分类[J].计算机工程与科学,2018, 40(06): 1125-1139.Z. Zhang, Q. Zou, Y. Lin, L. Chen and S. Wang.Improved Deep Hashing With Soft Pairwise Similarity for Multi-Label Image Retrieval[J].IEEE Transactions on Multimedia,2020,22(2):540-553.

致谢

非常的感谢教授我这门课程的程述立教授！程述立教授是我最尊敬的老师，也是我最敬爱的老师，他不仅在学术上给予我极大的帮助，更是在生活中给了我无限的鼓励和关爱。程述立教授在我学习期间给予了我最大的帮助。是他让我的论文写作得以顺利完成。也是他让我受益匪浅，感谢程教授这一年来对我的培养。同时，也要感谢我的舍友以及班中其他同学的帮助与支持。在我的学习生活中，总会遇到一些问题，大家相互帮助，共同解决问题。使我对理论知识有了更深层次的理解与应