

# Bilibili 视频获取及评论弹幕分析

## 视频获取

在bilibili上我们可以观看错位时空，链接 <https://www.bilibili.com/video/BV14z4y1s7s2/>

通过搜寻查找到了选集的json数据，根据其网址 <https://api.bilibili.com/x/player/pagelist?bvid=BV14z4y1s7s2&jsonp=jsonp> 进行request请求

```
{code: 0, message: "0", ttl: 1, data: [...]}

code: 0
data: [...]
  0: {cid: 285783437, page: 1, from: "vupload", part: "艾辰—【错位时空】",
duration: 205, vid: "", weblink: "",...}
    cid: 285783437
    dimension: {width: 1920, height: 1080, rotate: 0}
    duration: 205
    from: "vupload"
    page: 1
    part: "艾辰—【错位时空】"
    vid: ""
    weblink: ""

    message: "0"
    ttl: 1
```

根据其结构，是嵌套的字典，通过字典的键取值，如下部分是采集部分，分别解析出name,cid。

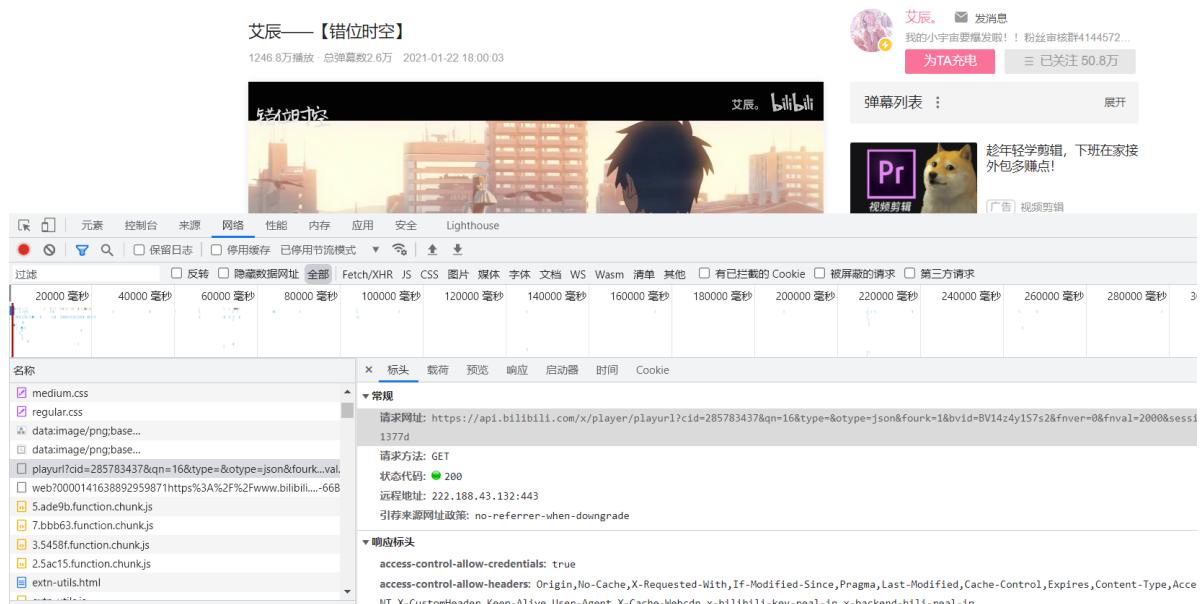
```
def get_cid_name(bvid):
    url = 'https://api.bilibili.com/x/player/pagelist?'
    param = {
        'bvid': bvid,
        'jsonp': 'jsonp'
    }
    response = requests.get(url=url, params=param, headers=headers).json()
    data = response['data'][0]

    cid = data['cid']
    name = data['part']
    return cid, name
```

```
cid = data['cid']  
name = data['part']
```

## 音视频链接搜寻

通过多次刷新，找到了音视频的真实链接，由下图此我们可以看出，想要获取视频链接就要先得到 bvid,cid,session三个值，以及一些其它固定的参数。cid、bvid已经获得，session则在视频html页面中直接存在，通过正则表达式即可获取 `session = re.findall('session': "(.*?)"', response.text)[0]`



请求时，应附带上如下参数



其返回的也是一个json结构，通过5层字典解析，可以得到真正的链接

```

▼{code: 0, message: "0", ttl: 1,...}
  ↳ 1 0
  ▼data: {from: "local", result: "suee", message: "", quality: 16, format: "mp4", timelength: 204095,...}
    ▶ accept_description: ["高清 1080P+", "高清 1080P", "高清 720P", "清晰 480P", "流畅 360P"]
    ▶ accept_format: "hdf1v2,f1v,f1v720,f1v480,mp4"
    ▶ 2 kept_quality: [112, 80, 64, 32, 16]
    ▼dash: {duration: 205, minBufferTime: 1.5, min_buffer_time: 1.5,...}
      ▶ audio: [{id: 30280,...}, {id: 30216,...}, {id: 30232,...}]
        ▶ 0: {id: 30280,...} 音频链接
        ▶ 1: {id: 30216,...}
        ▶ 2: {id: 30232,...}
      dolby: null
      duration: 205
      minBufferTime: 1.5
      3 _buffer_time: 1.5
    ▼vid: 4 [[id: 80,...], {id: 80,...}, {id: 64,...}, {id: 64,...}, {id: 32,...}, {id: 32,...}, {id: 16,...}, {id: 16,...}]
      ▶ 0: {id: 80,...}
        ▶ SegmentBase: {Initialization: "0-979", indexRange: "980-1503"}
        ▶ backupUrl: [...]
        ▶ backup_url: [...]
        bandw: 5 1600787
        baseUrl1: "https://xy120x209x53x179xy.mcdn.bilibili.video.cn:4483/upgcxcode/37/34/285783437/285783437_nb2-1-30080.m4s?e=ig8euxZM2rNbd1h0hNvNC8BqJIZNbfpXbEqxTEto8BT"
        base_url1: "https://xy120x209x53x179xy.mcdn.bilibili.video.cn:4483/upgcxcode/37/34/285783437/285783437_nb2-1-30080.m4s?e=ig8euxZM2rNbd1h0hNvNC8BqJIZNbfpXbEqxTEto8BT"
        codecid: 7
      ...
  ...

```

```

def get_video_url(bvid, cid, session):
    url = 'https://api.bilibili.com/x/player/playurl'
    params = {
        'cid': cid,
        'bvid': bvid,
        'qn': '0',
        'type': '',
        'otype': 'json',
        'fourk': '1',
        'fnver': '0',
        'fnval': '976',
        'session': session
    }
    response = requests.get(url=url, params=params, headers=headers).json()
    # json解析, 五层字典取出实际url
    audio_url = response['data']['dash']['audio'][0]['baseUrl']
    video_url = response['data']['dash']['video'][0]['baseUrl']
    print("音视频链接已得到" + "音频: " + audio_url + "视频: " + video_url)
    return audio_url, video_url

```

## 文件保存

此类音视频文件需要以二进制形式解析存储，在request时便需要以`.content`形式接收，保存时使用常规的文件保存方式即可，只需将后缀名改成对应的`.mp3` `.mp4`。

```

def save_video(name, audio_url, video_url):
    """保存数据 r.content获取二进制内容"""
    # 请求音频数据, 返回的数据二进制形式解析
    audio_content = get_response(audio_url).content
    print("开始下载" + name + "音频")
    # 保存音频文件
    with open(f"{name}.mp3", mode='wb') as af:
        af.write(audio_content)
    print(name + "音频保存完成")

    # 请求视频数据, 返回的数据二进制形式解析
    video_content = get_response(video_url).content
    print("开始下载" + name + "视频")
    # 保存视频文件
    with open(f"{name}.mp4", mode='wb') as vf:
        vf.write(video_content)
    print(name + "视频保存完成")

```

## 音视频合并

得到的音视频文件还需要合并才能成为一个完整的视频，否则手动对齐有点困难，这里使用的是 `ffmpeg`，下载链接 <https://github.com/BtbN/FFmpeg-Builds/releases>，下载后解压，将其下的bin目录添加到系统环境变量path中，可以在cmd中输入 `ffmpeg -version` 来检查是否添加成功，这样当我们使用命令时系统便可以识别 `ffmpeg` 并执行命令。

进行合并时，可以在本路径打开cmd输入命令合并

```
ffmpeg -i 艾辰--【错位时空】.mp4 -i 艾辰--【错位时空】.mp3 -acodec copy -vcodec copy A  
艾辰--【错位时空】.mp4
```

但既然我们已经使用了python，就贯彻到底，导包 `subprocess`，调用cmd命令

```
def merge_audio_video(name):  
    """音视频合并"""  
    # cmd 命令  
    command = f"ffmpeg.exe -i {name}.mp4 -i {name}.mp3 -acodec copy -vcodec copy  
A{name}.mp4"  
    subprocess.run(command, shell=True)  
    print(name + "视频正在执行合成命令")
```

至此，一个bilibili视频下载程序就完成了

## 完整程序

```
import pprint  
import re  
import subprocess  
import requests  
  
headers = {  
    # 浏览器标识符  
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36  
(KHTML, like Gecko) '  
    'Chrome/96.0.4664.45 Safari/537.36',  
    # 防盗链  
    'referer': 'https://www.bilibili.com/video/'  
}  
  
def get_response(url):  
    """发送请求"""  
    response = requests.get(url=url, headers=headers)  
    return response  
  
def get_cid_name(bvid):  
    url = 'https://api.bilibili.com/x/player/pagelist?'  
    param = {  
        'bvid': bvid,  
        'jsonp': 'jsonp'  
    }  
    response = requests.get(url=url, params=param, headers=headers).json()  
    data = response['data'][0]
```

```
cid = data['cid']
name = data['part']
return cid, name

def get_session(bvid):
    """request请求后 正则表达式session"""
    url = f'https://www.bilibili.com/video/{bvid}'
    response = requests.get(url=url, headers=headers)
    # 使用正则表达式进行获取
    session = re.findall('"session": "(.*?)"', response.text)[0]
    return session

def get_video_url(bvid, cid, session):
    url = 'https://api.bilibili.com/x/player/playurl'
    params = {
        'cid': cid,
        'bvid': bvid,
        'qn': '0',
        'type': '',
        'otype': 'json',
        'fourk': '1',
        'fnver': '0',
        'fnval': '976',
        'session': session
    }
    response = requests.get(url=url, params=params, headers=headers).json()
    # json解析, 五层字典取出实际url
    audio_url = response['data']['dash']['audio'][0]['baseUrl']
    video_url = response['data']['dash']['video'][0]['baseUrl']
    print("音视频链接已得到" + "音频: " + audio_url + "视频: " + video_url)
    return audio_url, video_url

def save_video(name, audio_url, video_url):
    """保存数据 r.content获取二进制内容"""
    # 请求音频数据, 返回的数据二进制形式解析
    audio_content = get_response(audio_url).content
    print("开始下载" + name + "音频")
    # 保存音频文件
    with open(f'{name}.mp3', mode='wb') as af:
        af.write(audio_content)
    print(name + "音频保存完成")

    # 请求视频数据, 返回的数据二进制形式解析
    video_content = get_response(video_url).content
    print("开始下载" + name + "视频")
    # 保存视频文件
    with open(f'{name}.mp4', mode='wb') as vf:
        vf.write(video_content)
    print(name + "视频保存完成")

def merge_audio_video(name):
    """音视频合并"""
    # cmd 命令
```

```

command = f"ffmpeg.exe -i {name}.mp4 -i {name}.mp3 -acodec copy -vcodec copy
A{name}.mp4"
subprocess.run(command, shell=True)
print(name + "视频正在执行合成命令")

if __name__ == '__main__':
    bvid = 'BV14z4y1S7s2'
    # 调用选集获取函数，信息填入三个列表
    cid, name = get_cid_name(bvid)
    # 获取session
    session = get_session(bvid)
    # 通过前文搜集到信息，寻找音视频真正的链接
    audio_url, video_url = get_video_url(bvid, cid, session)
    # 根据音频，视频链接分别下载
    save_video(name, audio_url, video_url)
    # 使用ffmpeg.exe合并下载到的音视频
    merge_audio_video(name)

```

## 控制台输出结果正常

```

285783437 艾辰—【错位时空】
音视频链接已得到音频: https://xy120x209x53x176xy.mcdn.bilibili.com:4483/upgcxcode/37/34/285783437/285783437\_nb2-1-30280
.m4s?e
=ig8euxZM2rNcNbd1hoNvNC8BqJ1zNbfqXByEqxTEto8BTnVnN0GvT90W5JZMKx_YN0MvXg8gNEV4NC8xNEV4N03eN0B5tZlqNxTEto8BTnVnNeZvJ10K1_g2UB02J0mN0B5tZlqNCNEto8BTnVnC7MTX502C8f2jM0J6mqF2f
ka1mqx6qqj0eN0B59M-&upk=56nbs=16deadline=1638902730&gen=playurlv2&os=mcdn&oi=2025934354&trid=0001b1ac67f14a1a4c579579b51d2a4956feu&platform=pc&upsid
=37b6bce72ce1de39b1bf1b029b402d40&uparams=e,upk,nbs,deadline,gen,os,oi,trid,platform&mcnid=8000408&mid=0&bw=vod&nettype=0&order_id=0,3&agr=1&bw=39911&logo=A0000080视频,
https://xy120x209x53x176xy.mcdn.bilibili.com:4483/upgcxcode/37/34/285783437/285783437\_nb2-1-30080
.m4s?e
=ig8euxZM2rNcNbd1hoNvNC8BqJ1zNbfqXByEqxTEto8BTnVnN0GvT90W5JZMKx_YN0MvXg8gNEV4NC8xNEV4N03eN0B5tZlqNxTEto8BTnVnNeZvJ10K1_g2UB02J0mN0B5tZlqNCNEto8BTnVnC7MTX502C8f2jM0J6mqF2f
ka1mqx6qqj0eN0B59M-&upk=56nbs=16deadline=1638902730&gen=playurlv2&os=mcdn&oi=2025934354&trid=0001b1ac67f14a1a4c579579b51d2a4956feu&platform=pc&upsid
=187298bae64183a79f3e614eedc71a1b6&uparams=e,upk,nbs,deadline,gen,os,oi,trid,platform&mcnid=8000410&mid=0&bw=vod&nettype=0&order_id=0,3&agr=1&bw=2000098&logo=A0000080
开始下载艾辰—【错位时空】音频
艾辰—【错位时空】音频保存完成
开始下载艾辰—【错位时空】视频
艾辰—【错位时空】视频保存完成
ffmpeg version N-104708-g8a18db3ec6-20211201 Copyright (c) 2000-2021 the FFmpeg developers
  built with gcc 10-win32 (GCC) 20210610
  configuration: --prefix=ffmpeg/build --prefix= --pkg-config=pkg-config --cross-prefix=x86_64-w64-mingw32 --arch=x86_64 --target-os=mingw32
  -enable-version3 --disable-debug --disable-w32threads --enable-pthreads --enable-iconv --enable-libxml2 --enable-zlib --enable-libfreetype --enable-libfribidi --enable-gmp
  -enable-lzma --enable-fontconfig --enable-libvorbis --enable-opencl --enable-libvmaf --disable-libxcb --disable-xlib --enable-amf --enable-libaom --disable-avisynth
  --enable-libdavid --disable-libdav2s --disable-libfdk-aac --enable-ffnvenc --enable-cuda-llm --disable-frei0r --enable-libaom --enable-libass --enable-libbluray

```

## 音视频下载正常



A艾辰—【错位时空】.mp4



艾辰—【错位时空】.mp3



艾辰—【错位时空】.mp4

## 弹幕获取

通过查找找到了弹幕的xml文件 <https://comment.bilibili.com/285783437.xml>，其中的数字就是我们上面得到的额cid。

```
<!--<br><chatserver>chat.bilibili.com</chatserver>
<chatid>285783437</chatid>
<mission>0</mission>
<maxlimit>1000</maxlimit>
<state>0</state>
<real_name>real_name</real_name>
<source>kv</source>
<id>129_52090_1,25_16777215_1633449770_0,d42a3>29_58232962025007104,10</>再见了，仇信琳，这回是真的再见了，再也不见</d>
<id p="117_52090_1,25_16777215_1633449585_0,d42a3>29_58232965093075456,10">原来人真的是可以一瞬间忘记一段感情的</d>
<id p="195_10000_1,25_16765698_1638419645_0,52>67ha0_5826817817624576,10</>两袖清风，怎教佳人入！</d>
<id p="110_28400_1,25_38979_1632183043_0,460a92a_58244120193675776,10</>一身清贫怎融入繁华，两袖清风岂敢误佳人</d>
<id p="151_51400_5,25_52480_1632111887_0,d42a2e>21_58206814249707200,10</>“告我啊，求你了”</d>
<id p="138_86200_5,25_52480_1632111799_0,d42a2e>21_5820676748705000,10</>每日一遭，欢迎大不列颠意识体来我们当唱见<脱单doge></d>
<id p="165_62800_5,25_52480_1632111894_0,ff140c57_58018074262190086,10</>我要的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="155_36700_1,25_16777215_165775187_0,ff140c57_58018062300573164,10</>我要的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="105_65400_1,25_16777215_1633449585_0,d42a3>29_58232965093075456,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界</d>
<id p="59_36700_5,25_38979_1637489399_0,6e99a736_5720045102516674723,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉，那这个世界应该会对同性恋宽容些吧</d>
<id p="92_26800_5,25_38979_1637420497_0,1e79a497_5724422657945723,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="113_49700_5,25_52480_1637149037_0,13a14f72_5770194093125248,10</>我要的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="49_06500_1,25_16777215_163617046_0,c0258a91_57183947877649408,10</>爱情防控，算！</d>
<id p="167_14500_5,25_1513823_1635263490_0,2b4d9ect_5671343387320704,10</>从此山水不相离，莫问旧人长与短</d>
<id p="89_81000_5,25_38979_1634383702_0,50a3>3ce_56252171562299248,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="75_71300_1,25_41904_1634355595_0,ff2a2d4_56252171562299248,10</>致尊敬的宋老师，我已黑暗，仰望你的皎洁</d>
<id p="105_65400_1,25_16777215_1633449585_0,ff2a2d4_56252171562299248,10</>他用魔法吟唱着唤起奇迹的咒语，无论何时都熠熠生辉</d>
<id p="13_41900_1,25_38979_1633282777_0,29<0>08620_55237491900935712,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界</d>
<id p="18_41900_1,25_38979_16328592772_0,29<0>08620_55237491900935712,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界</d>
<id p="12_09900_1,25_38979_16325592766_0,29<0>08620_55237491900935712,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界</d>
<id p="0_77900_1,25_16777215_1633337468_0,38469712_557036435453568,10</>“世界上最难过的的事情，大概就是喜欢上纸片人了吧。”</d>
<id p="38_00000_1,25_16777215_1633260563_0,f9674709_55663323194763776,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="38_96700_5,25_38979_1633233176_0,ace63c88_5564936740795648,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉！</d>
<id p="63_07300_1,25_38979_1633073377_0,7542107a_5556544594843072,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界。</d>
<id p="199_24500_5,25_52480_1632055375_0,7e9fe2d5_555557452700800,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="187_27600_1,25_52480_1632055363_0,7e9fe2d5_555557395473686,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="180_21200_1,25_52480_1632055365_0,7e9fe2d5_5555573549520128,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="179_21200_1,25_52480_1632055364_0,7e9fe2d5_5555573549520128,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="160_72800_1,25_52480_1632055324_0,7e9fe2d5_555557254342528,10</>我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉</d>
<id p="70_43200_5,25_52480_1622935508_0,2e4dab14_5547087925377152,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界。</d>
<id p="49_37200_1,25_16777215_1622392494_0,2e4dab14_5547088209532560,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界。</d>
<id p="101_21800_5,25_48128_1632824344_0,ab6d1714_55434618953407488,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界。</d>
<id p="33_02200_5,25_38979_1622221900_0,1663450b_55118764693528064,10</>他用魔法吟唱着唤起奇迹的咒语，带着红茶与玫瑰出现在这个世界。</d>
```

对于如此整齐的页面，我们可以使用正则表达式 `<d p=". *?">(. *?)</d>` 得到一个弹幕列表

```
import re
import requests

url = 'https://comment.bilibili.com/285783437.xml'
# 根据网址请求
response = requests.get(url=url)
# 使用 utf-8 编码
response.encoding = 'utf-8'
# 将得到的响应解析为文本
text = response.text

comments = re.findall('<d p=". *?">(. *?)</d>', text)
print(comments)

with open("弹幕.txt", mode='a', encoding='utf-8') as f:
    for comment in comments:
        f.write(comment + '\n')
        print(comment)
```

通过这种方式获得了1059条弹幕

评论.txt - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉  
负累啊，让我抛弃你吧  
我吹过你吹过的晚风  
不要喜欢别人好不好，为什么不能喜欢我  
洪佳壕我爱你  
我吹过你吹过的晚风  
陈萍萍，你在看吗，我c爱你一辈子  
贾斯喻，再见了  
妈妈，我想你了，快回来吧  
我吹过你吹过的牛逼  
我以为我们最起码是朋友，可早已经不是了  
张怡，我爱你！！！！！！不要离开  
\\\\\\\\\  
LXY,我想你了  
.  
蒋芸蕊 我爱你  
宸，希望你记得我  
翾，你一定要幸福啊  
我爱的人有一双祖母绿的眼睛，无论何时都熠熠生辉！！！！！！  
泰晤士河上空的残阳是他他荣光不散的余晖！！！！！  
我爱的人有一双祖母绿的眼眸，无论何时都熠熠生辉  
我爱的人有一双祖母绿的眼眸，无论何时都熠熠生辉  
给时间一点时间，让过去的过去，让开始的开始！！！！！！！  
我喝过你喝过的西北风，那我算不算比你穷

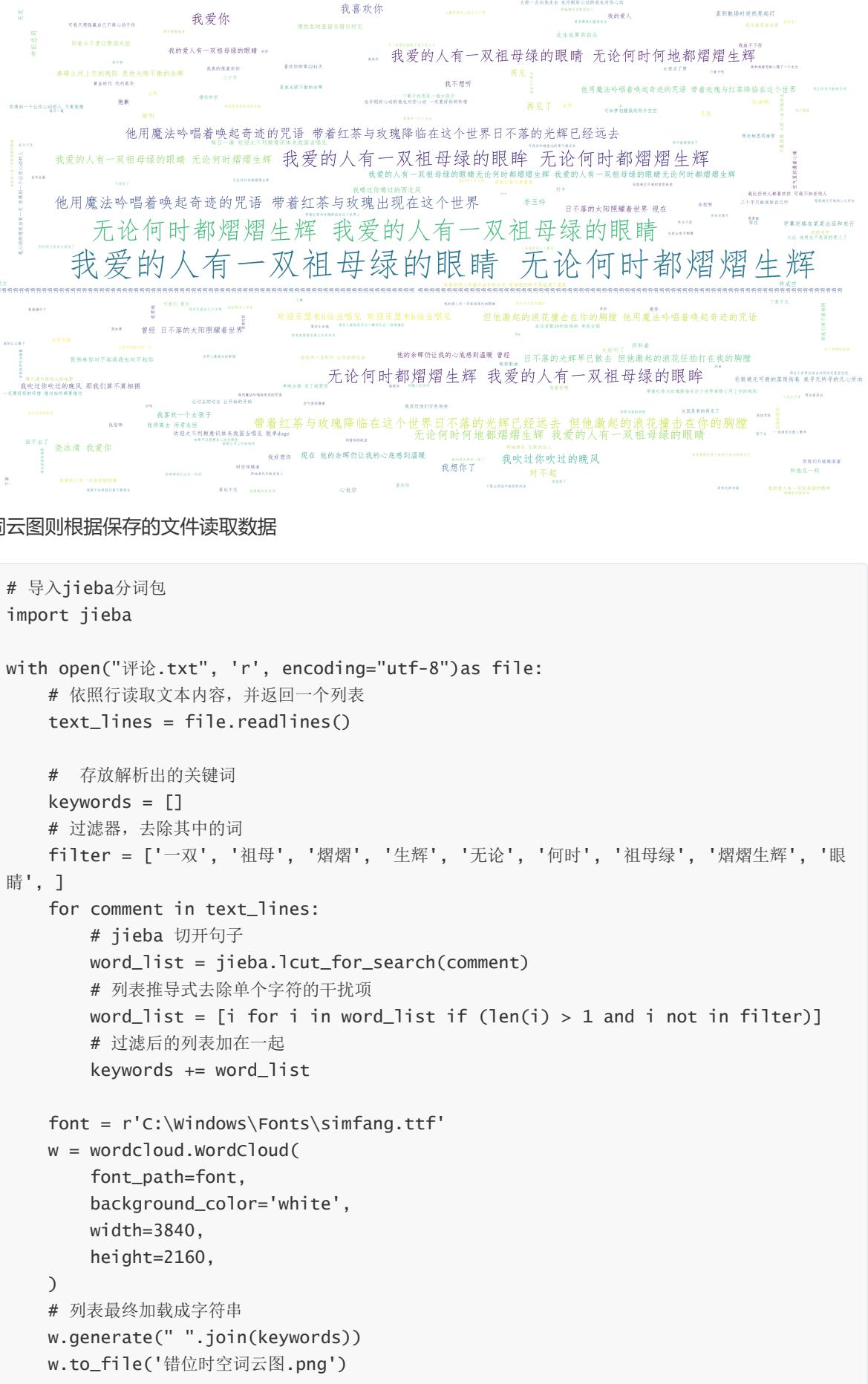
第 1059 行, 第 20 列 100% Windows (CRLF) UTF-8

## 弹幕云图

将弹幕转为云图是一种较好的可视化方式

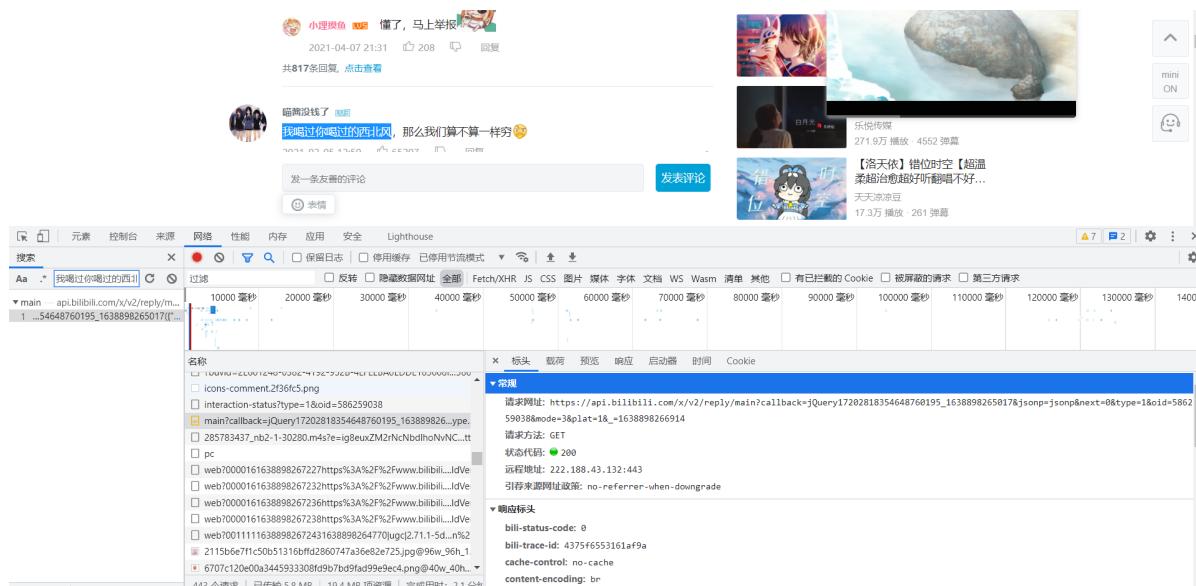
句云图直接从上面程序得到的遍历中取出

```
# 遍历评论
sentence = ""
for comment in comments:
    # print(comment)
    sentence = sentence + ' ' + comment
font = r'C:\Windows\Fonts\simfang.ttf'
w = wordcloud.WordCloud(
    font_path=font,
    background_color='white',
    width=3840,
    height=2160,
)
w.generate(sentence)
w.to_file('错位时空句云图.png')
```





## 评论获取



根据搜索我们直接找到了其链接，及请求时所需要的参数

× 标头 载荷 预览 响应 启动器 时间 Cookie

▼ 查询字符串参数      查看来源      查看网址编码格式的数据

```

callback: jQuery17202818354648760195_1638898265017
jsonp: jsonp
next: 0
type: 1
oid: 586259038
mode: 3
plat: 1
_: 1638898266914

```

其所返回的同样是json类型数据，评论信息也存在于其中

× 标头 载荷 预览 响应 启动器 时间 Cookie

```

cm: {}
► cm_info: {ads: {4765: [...]}}
► config: {showadmin: 1, showentry: 1, showfloor: 0, showtopic: 1, show_up_flag: true, read_only: false,...}
► control: {input_disable: false, root_input_text: "发一条友善的评论", child_input_text: "",...}
► cursor: {all_count: 16102, is_begin: true, prev: 1, next: 2, is_end: false, mode: 3, show_type: 1,...}
► effects: {preloading: ""}
► folder: {has_folded: false, is_folded: false, rule: "https://www.bilibili.com/blackboard/foldingreply.html"}
hots: null
lottery: 0
lottery_card: null
note: 1
notice: null
► replies: [...]
  ▼ 0: {rid: 4061055432, oid: 586259038, type: 1, mid: 598648539, root: 0, parent: 0, dialog: 0, count: 628,...}
    action: 0
    assist: 0
    attr: 0
    ▼ content: {message: "我喝过你喝过的西北风，那么我们算不算一样穷[思考]", plat: 0, device: "", members: [],...}
      device: ""
      ► emote: {[思考]: {id: 1941, package_id: 1, state: 0, type: 1, attr: 0, text: "[思考]",...}}
      jump_url: {}
      max_line: 6
      members: []
      message: "我喝过你喝过的西北风，那么我们算不算一样穷[思考]"
      plat: 0
    count: 628

```

此处的有效信息均位于replies键对应的值中，可采取分步解析

```

reply = response['data']['replies'][i]
# 将评论时间的时间戳格式转为标准时间格式
t = reply['ctime']
timeArray = time.localtime(t)
otherStyleTime = time.strftime("%Y-%m-%d %H:%M:%S", timeArray)
comment_dict = {
    "评论": reply['content']['message'].replace("\n", ""),
    '时间': otherStyleTime,
    '点赞数': reply['like']
}

```

通过选出两个请求链接比对，以及手动输入进浏览器试探，最终发现第一个参数与最后一个参数无用（不影响输出，而且加上反而无法请求），next参数类似于翻页。oid为该视频的标识符。mode为评论的排序方式，2表示按时间排序，3表示按热度排序。

```
https://api.bilibili.com/x/v2/reply/main?  
callback=jQuery17202818354648760195_1638898265017&jsonp=jsonp&next=0&type=1&oid=586259038&mode=3&plat=1_=1638898266914  
https://api.bilibili.com/x/v2/reply/main?  
callback=jQuery17202818354648760195_1638898265019&jsonp=jsonp&next=3&type=1&oid=586259038&mode=3&plat=1_=1638899174477
```

得到的有效api为 <https://api.bilibili.com/x/v2/reply/main?jsonp=jsonp&next=0&type=1&oid=208143004&mode=3&plat=1>

评论获取程序

```
import csv  
import time  
import requests  
  
headers = {  
    # 浏览器标识符  
    'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) '  
    'Chrome/96.0.4664.45 Safari/537.36',  
    # 防盗链  
    'referer': 'https://www.bilibili.com/video/'  
}  
  
  
def get_response(url):  
    """发送请求"""  
    response = requests.get(url=url, headers=headers).json()  
    return response  
  
  
def parse_response(response):  
    """对得到的json数据进行解析, 选出评论"""  
  
    for i in range(20):  
        try:  
            reply = response['data']['replies'][i]  
            # 将评论时间的时间戳格式转为标准时间格式  
            t = reply['ctime']  
            timeArray = time.localtime(t)  
            otherStyleTime = time.strftime("%Y-%m-%d %H:%M:%S", timeArray)  
            comment_dict = {  
                "评论": reply['content']['message'].replace("\n", ""),  
                '时间': otherStyleTime,  
                '点赞数': reply['like']  
            }  
            fp.writerow(comment_dict)  
        except Exception as e:  
            print("本评论解析写入失败" + str(i) + str(e))  
  
  
if __name__ == '__main__':  
    oid = 586259038  
  
    with open("评论.csv", mode="a", encoding='utf-8', newline="") as f:
```

```

fp = csv.DictWriter(f, fieldnames=['评论', '时间', '点赞数'])
fp.writeheader() # 写入表头
for next in range(10):
    url = f'https://api.bilibili.com/x/v2/reply/main?jsonp=jsonp&next={next}&type=1&oid={oid}&mode=3&plat=1'
    response = get_response(url)
    parse_response(response)

```

共获得了199条评论数据

```

181  这么大一个up个王丁六孩都叫个吧,2021-07-11 13:21:11,38
182  艾辰黑历史合集,均已石锤1.原头像描图亚瑟柯克兰,态度极其恶劣,且并不是全然无知,17年用原图做过头像2.视频《痒》直接盗用aph的图(小菊亚瑟)【艾辰】痒(受害者)
183  他用魔法吟唱着唤起奇怪的咒语,带着红茶与玫瑰降临在这个世界泰晤士河上空的残阳,是他光荣不散的余晖,2021-07-17 10:43:24,37
184  艾辰黑历史:1原头像描图亚瑟柯克兰,态度极其恶劣2视频直接盗用aph的图(小菊亚瑟)3无版权翻唱撒野还魔改4镜头像人设是偷的零度恋人的,来人细说5《念》Pv是描冰果的6网易云?
185  《关于UK转生到bilibili被唱见这件事》,2021-07-15 00:23:06,37
186  看大家一直在这里骂他也没什么意思,不如我们学着老福特的样子把这一变成扩列楼得了【藏狐】,2021-07-14 20:03:48,37
187  万万没想到,此人竟然是唱见圈xz,2021-07-14 10:29:01,37
188  请问宁是哪块殖民地【doge】,2021-07-14 17:28:11,60
189  晚上好,唱见圈xz,请尽早糊掉地心噢【给心心】,2021-08-13 00:08:30,47
190  看到弹幕,全是懦弱小学生告白稿,笑死,难怪全是悲剧,2021-09-04 01:07:45,42
191  晚上好,唱见圈xz,每日礼貌一问,请问宁糊了吗之前评论有个姐妹说宁道歉了让我删评,我删了,但最近又看见了宁的瓜,看了宁还是一如既往的恶心啊【给心心】希望宁早日凉透糊穿宇宙
192  生活不易,UK卖艺,2021-07-15 00:21:31,36
193  来等道歉的朋友记得点个踩电弹幕提醒再走【脱单doge】【脱单doge】【脱单doge】我顺便发个祝福:ac与xz百年好合。描图还骂番的人不配哈【脱单doge】,2021-07-14:
194  “纯路人,第一次听艾辰唱歌的时候是牵丝戏,那个时候我就觉得这个人唱歌好好听!然后一下就记住了。但一次也没怎么关注,今天就刷到了,点进来觉得还是那么好听(歌声真的太绝了)
195  恶心向我们英厨道歉以及和画手道歉【微笑】,2021-07-13 14:34:12,41
196  歉不错。人不行。,2021-08-13 13:03:33,35
197  懊愧,听了几个月五四特别版才知道原来是翻唱。慕名而来听了原曲,实话实说,原词真挺口水的,青春期的填词境界就高了一档,尤其是配合觉醒年代mv或者奥运mv,搬上春晚都不为过
198  好吧,断断续续的了解了下。当初我用B站只看番时,就从音乐软件听过他歌了,也算是粉丝(听歌的那种)。就是,没想到音乐区的up能搞出这么多事。取关献上,2021-07-20 00:47
199  (转)姐妹们姐妹们,现在艾辰可能不会管b站这边了,艾辰昨天还在dy发布了歌,我们现在应该把战力转移到dy那边,抖音现在姐妹少的可怜【时光代理人_来活了】,2021-07-19 14:47
200  那个艾辰有没有为他的粉丝踢过欧洲杯,为他的粉丝,卖点什么的,创造点一些np点的科技。你们艾辰有没有给你们发明过汽车还创作官方谷子。有没有为你们美帝瞎过眼啥的,唉,没关

```

## 评论分析

对获取的199条评论进行分析,统计最初的几十天评论的点赞数,并绘制出折线图。

```

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

# 读入评论数据 并分行
df = pd.read_csv("评论.csv")
df.columns = ['评论', '时间', '点赞数']

# 为了使分月时,第二列显示出数量重新分列,
df.columns = ['评论', '时间', '点赞数']
# 将数据类型转换为日期类型
df['时间'] = pd.to_datetime(df['时间'])
# 将date设置为index
df_time = df.set_index('时间')
# 根据天进行分离
df_month = df_time.resample('D').sum().to_period('D')
# 截取前26天的评论点赞数,间接反应了当时人气
df_month = df_month[0:26]

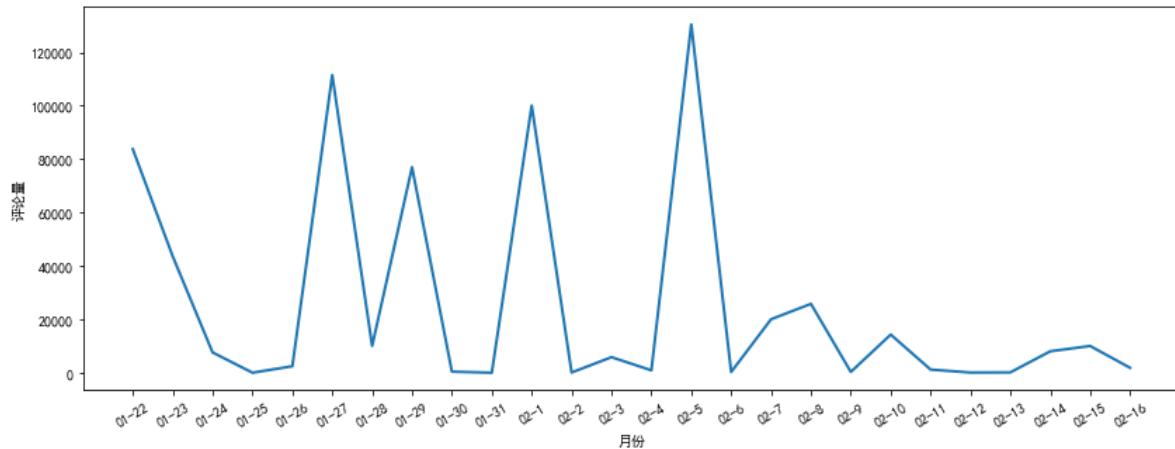
#设置中文显示
plt.rcParams['font.sans-serif'] = 'SimHei'
# make data
day_list = []
for d in range(22, 32):
    day_list.append(f"01-{d}")
for d in range(1, 17):
    day_list.append(f"02-{d}")

```

```

# plot
fig, ax = plt.subplots(figsize=(14,5))
ax.plot(day_list,df_month['点赞数'], linewidth=2.0)
plt.xticks(rotation=30)
ax.set_xlabel("月份")
ax.set_ylabel("评论量")
plt.show()

```



继续不忘初心生成词云图

```

# 导入jieba分词包
import jieba
import wordcloud

# 存放解析出的关键词
keywords = []
# 过滤器, 去除其中的词
filter = ['呜呜', '亚瑟', '克兰', '不算']
for comment in df['评论']:
    # jieba 切开句子
    word_list = jieba.lcut_for_search(comment)
    # 列表推导式去除单个字符的干扰项
    word_list = [i for i in word_list if (len(i) > 1 and i not in filter)]
    # 过滤后的列表加在一起
    keywords += word_list

font = r'C:\Windows\Fonts\simfang.ttf'
w = wordcloud.WordCloud(
    font_path=font,
    background_color='white',
    width=3840,
    height=2160,
)
# 列表最终加载成字符串
w.generate(" ".join(keywords))
w.to_file('错位时空评论词云图.png')

```

