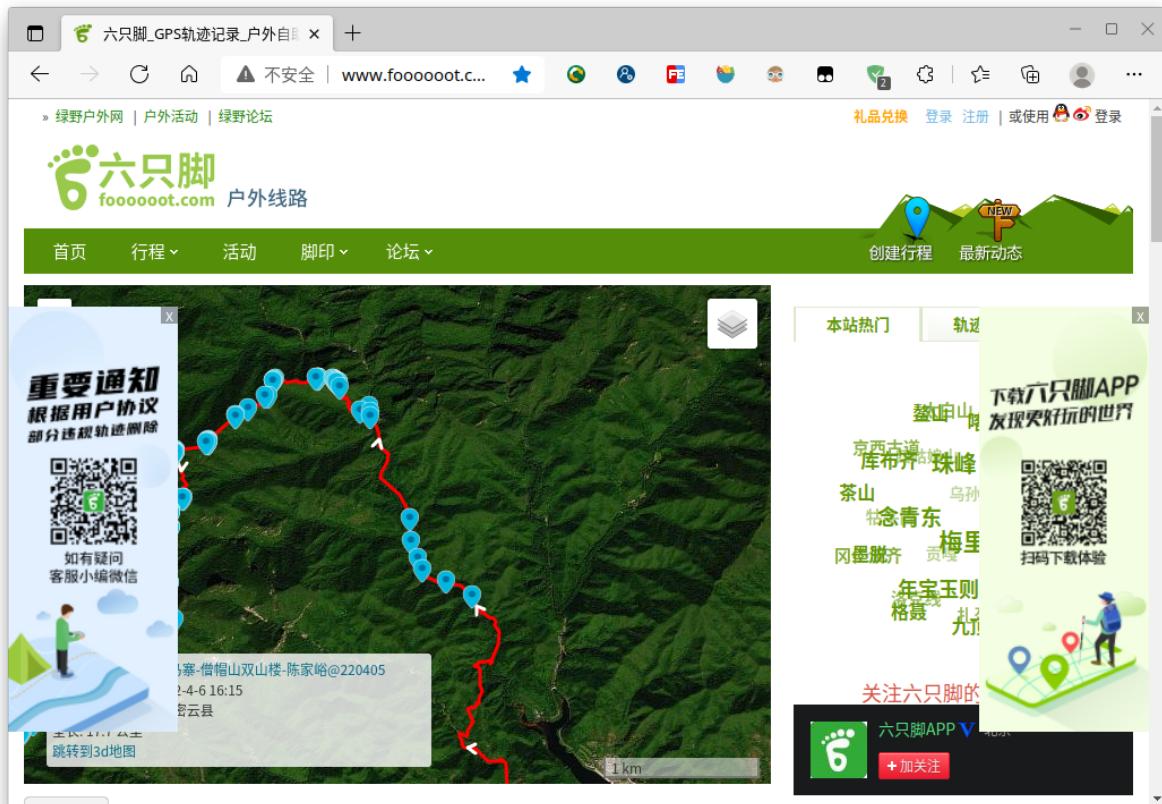


# 1. 引言

六只脚是国内著名的户外网站，拥有大量的户外GPS轨迹路线，网址为：

<http://www.fooooooot.com/>



# 2. 数据分析

## 2.1 获取所有轨迹

搜索关键词 岳麓山：

中国\_行程\_GPS轨迹\_GPS线路

六只脚 foooooot.com 户外线路

类型: 全部 观光旅行 徒步 登山 骑行 自驾车 跑步 摄影 越野跑 Gps作画 山地骑行 摩托车 寻宝 滑雪 航海 滑翔伞 滑板 其他 导入

地域: 境外 境内 北京 天津 河北 山西 内蒙古 辽宁 吉林 黑龙江 上海 江苏 浙江 安徽 福建 江西 山东 河南 湖北 湖南 广东 广西 海南 重庆 四川 贵州 云南 西藏 陕西 甘肃 青海 宁夏 新疆 台湾 香港 澳门

关键字: 岳麓山 搜索

显示推荐行程 显示带图片行程 排序: 默认排序 在地图上查看

**岳麓山 (岳麓书院)**  
SS铁驴 2017-11-22上传 徒步 8.7公里 难度: 简单  
于2017-11-22 13:36 出发,历时 4 hours, 44 minutes  
湖南 长沙 长沙市区 - 湖南长沙 长沙市区

**岳麓山, 岳麓书院**  
苏格拉底的业余粉丝 2016-3-14上传 观光旅行 8.2公里 难度: 简单  
于2016-03-14 14:44 出发,历时 1 hour, 56 minutes  
湖南 长沙 长沙市区 - 湖南长沙 长沙市区

**岳麓山桃花岭岳麓山**  
bovong@qq 2015-1-19上传 越野跑 29.8公里 难度: 简单

看看TA们都去了哪

可以发现，每页具有三十个轨迹记录

观察第二页的网址我们可以发现网址为：

<http://www.foooooot.com/search/trip/all/1/all/time/descent/?page=2&keyword=%E5%B2%B3%E9%BA%93%E5%B1%B1> 不难发现其规律：

- page为页数
- keyword为 岳麓山 的转义

我们不妨测试page为50的情况，在浏览器输入

<http://www.foooooot.com/search/trip/all/1/all/time/descent/?page=50&keyword=%E5%B2%B3%E9%BA%93%E5%B1%B1> :

中国\_行程\_GPS轨迹\_GPS线路

绿野户外网 | 户外活动 | 绿野论坛

礼品兑换 登录 注册 | 或使用

六只脚 fooot.com 户外线路

首页 行程 活动 脚印 论坛

创建行程 最新动态

类型: 全部 观光旅行 徒步 登山 骑行 自驾车 跑步 摄影 越野跑 Gps作画 山地骑行 摩托车 寻宝 滑雪 航海 滑翔伞 滑板 其他 导入

地域: 境外 境内 北京 天津 河北 山西 内蒙古 辽宁 吉林 黑龙江 上海 江苏 浙江 安徽 福建 江西 山东 河南 湖北 湖南 广东 广西 海南 重庆 四川 贵州 云南 西藏 陕西 甘肃 青海 宁夏 新疆 台湾 香港 澳门

关键字: 岳麓山 搜索

显示推荐行程 显示带图片行程 排序: 默认排序 在地图上查看

没有相关行程!

« 上一页 1 2 3 4 5 6 下一页 »

看看TA们都去了哪

香八拉往返徒步--公园一北京西山,国家森林公园--公园。--西郊线--站一巴沟 2022-04-14

岳麓山

小结：我们可以通过不断增加page的数字，直至某一页不满足三十个轨迹记录，获取该关键词所有的轨迹记录

## 2.2 获取轨迹ID

点击某个具体的轨迹详情：

中国\_行程\_GPS轨迹\_GPS线路

岳麓山（岳麓书院）\_行程\_6

当前位置: 首页 -> 全部行程 -> 中国 - 湖南省 - 长沙市 - 岳麓山（岳麓书院）

六只脚 fooot.com 户外线路

首页 行程 活动 脚印 论坛

创建行程 最新动态

岳麓山（岳麓书院） SS铁驴 2017-11-22 18:23 湖南长沙长沙市区到湖南长沙长沙市区 全长: 8.7公里 带了岭 跳转到3d地图

SS铁驴 2017-11-22 18:23 湖南长沙长沙市区到湖南长沙长沙市区 全长: 8.7公里 带了岭 跳转到3d地图

评论 0 次, 下载 2 次, 收藏 6 次, 浏览 825 次

同行驴友(共 1 名)

SS铁驴

下载GPS轨迹文件

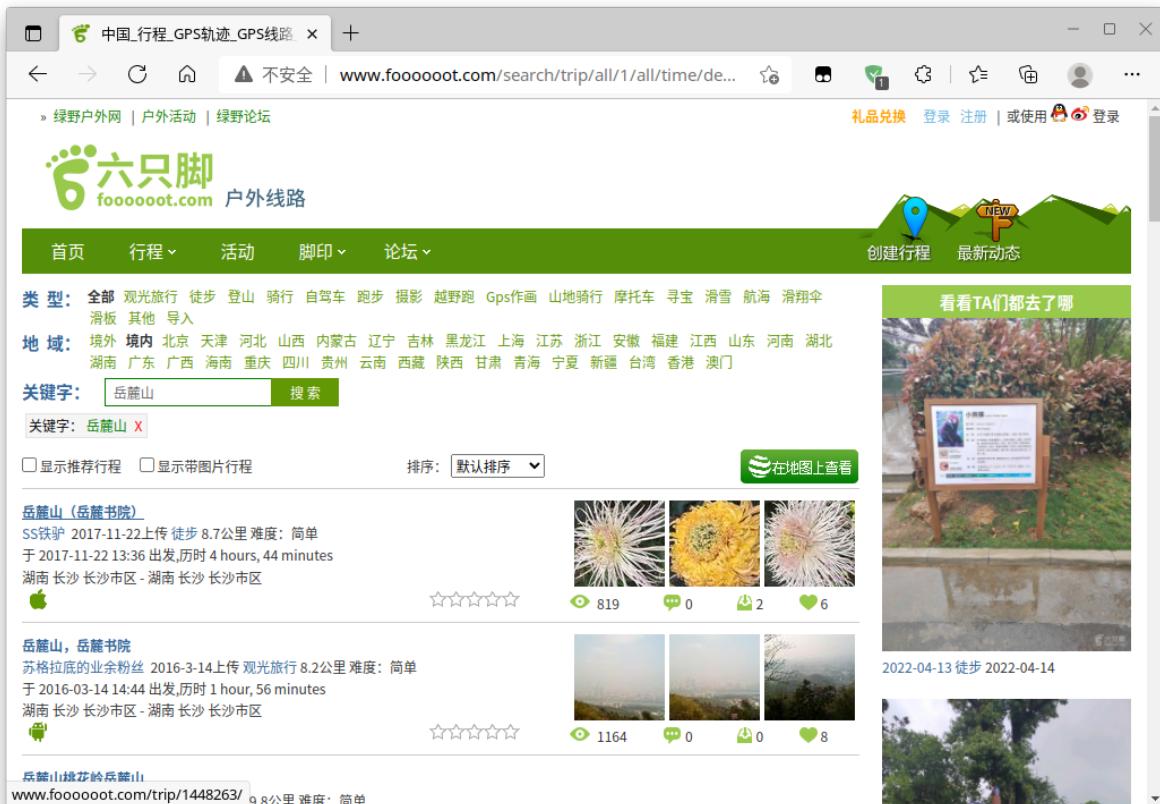
下载GPS文件 提供作者原始GPS轨迹、航点文件下载及 GPX轨迹记录、KML轨迹文件下载。

所用设备

可以看到每一页具体的轨迹页面的网址是由轨迹ID构造的，诸如：

<http://www.fooooooot.com/trip/1448263/>

从刚才的列表界面我们就可以找到每个轨迹ID：



中国\_行程\_GPS轨迹\_GPS线路

六只脚 fooooooot.com 户外线路

类型：全部 观光旅行 徒步 登山 骑行 自行车 跑步 摄影 越野跑 Gps作画 山地骑行 摩托车 寻宝 滑雪 航海 滑翔伞 滑板 其他 导入

地域：境内 北京 天津 河北 山西 内蒙古 辽宁 吉林 黑龙江 上海 江苏 浙江 安徽 福建 江西 山东 河南 湖北 湖南 广东 广西 海南 重庆 四川 贵州 云南 西藏 陕西 甘肃 青海 宁夏 新疆 台湾 香港 澳门

关键字：岳麓山 搜索

岳麓山 (岳麓书院)  
SS铁驴 2017-11-22上传 徒步 8.7公里 难度：简单  
于2017-11-22 13:36 出发,历时 4 hours, 44 minutes  
湖南 长沙 长沙市 - 湖南长沙 长沙市区

岳麓山, 岳麓书院  
苏格拉底的业余粉丝 2016-3-14上传 观光旅行 8.2公里 难度：简单  
于2016-03-14 14:44 出发,历时 1 hour, 56 minutes  
湖南 长沙 长沙市 - 湖南长沙 长沙市区

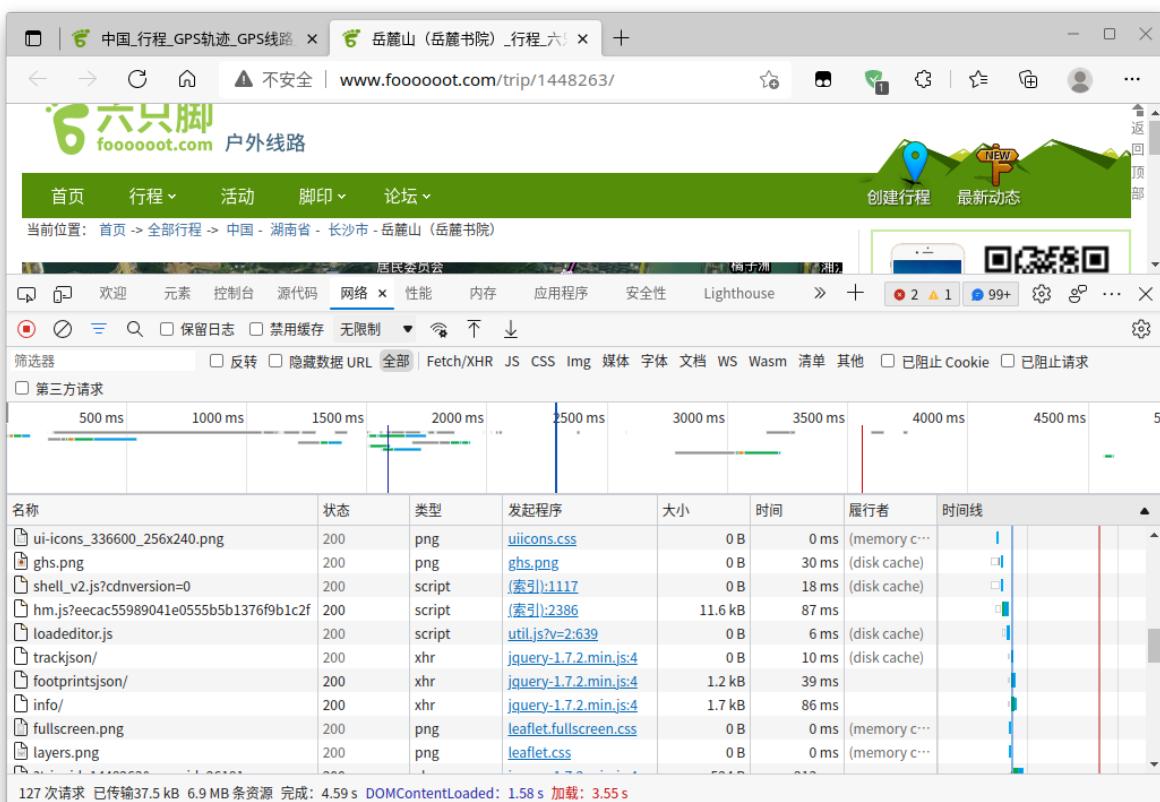
岳麓山桃花岭岳麓山  
www.fooooooot.com/trip/1448263/ 9.8公里 难度：简单

看看TA们都去了哪

在地图上查看

## 2.3 获取轨迹数据

我们打开浏览器控制台（按F12），点击到网络记录界面，刷新网址：



中国\_行程\_GPS轨迹\_GPS线路

岳麓山 (岳麓书院) 行程 六

六只脚 fooooooot.com 户外线路

居民委员会

500 ms 1000 ms 1500 ms 2000 ms 2500 ms 3000 ms 3500 ms 4000 ms 4500 ms 5

名称	状态	类型	发起程序	大小	时间	履行者	时间线
ui-icons_336600_256x240.png	200	png	uiicons.css	0 B	0 ms	(memory c...)	
ghs.png	200	png	ghs.png	0 B	30 ms	(disk cache)	
shell_v2.js?cdnversion=0	200	script	[索引]:1117	0 B	18 ms	(disk cache)	
hm.js?eeac55989041e0555b5b1376f9b1c2f	200	script	[索引]:2386	11.6 kB	87 ms		
loadeditor.js	200	script	util.js?v=2:639	0 B	6 ms	(disk cache)	
trackjson/	200	xhr	jquery-1.7.2.min.js:4	0 B	10 ms	(disk cache)	
footprintsjson/	200	xhr	jquery-1.7.2.min.js:4	1.2 kB	39 ms		
info/	200	xhr	jquery-1.7.2.min.js:4	1.7 kB	86 ms		
fullscreen.png	200	png	leaflet.fullscreen.css	0 B	0 ms	(memory c...)	
layers.png	200	png	leaflet.css	0 B	0 ms	(memory c...)	

127 次请求 已传输37.5 kB 6.9 MB 资源 完成: 4.59 s DOMContentLoaded: 1.58 s 加载: 3.55 s

从网络请求记录中我们发现有两个XHR异步请求其名字很像轨迹数据，点开查看：

The screenshot shows the Network tab in the Chrome DevTools. There are two main entries in the list:

- trackjson/**: This entry has a expanded dropdown showing 12 items, each containing a timestamp and coordinates (e.g., [1511329003, 28.171333, 112.940339]).
- footprintsjson/**: This entry also has a expanded dropdown showing 12 items, each containing a timestamp and coordinates.

Both requests have a response size of 37.5 kB and a duration of 6.9 MB.

可以看到，这个trackjson就是轨迹的JSON数据：

The screenshot shows the Network tab in the Chrome DevTools, identical to the one above. It displays the same two XHR requests for 'trackjson/' and 'footprintsjson/'. The 'trackjson/' request is expanded to show 12 items of trajectory data, and the 'footprintsjson/' request is expanded to show 12 items of footprint data.

这个footprintsjson就是足迹数据，也就是拍照的那种数据：

The screenshot shows a dual-pane NetworkMiner interface. The left pane displays a list of captured network requests, including file downloads and JSON data. The right pane shows detailed information for a selected request, including headers, preview, response, and cookie data. The browser's address bar shows a trip page from www.fooooooot.com.

名称	标头	预览	响应	发起程序	计时	Cookie
ui-icons_336600_256x240.pr						
ghs.png						
shell_v2.js?cdnversion=0						
hm.js?ecac55989041e0555...						
loadeditor.js						
trackjson/						
footprints.json/						
info/						
fullscreen.png						
layers.png						

127 次请求 已传输 37.5 KB 6.9 MB

## 2.4 字段解释

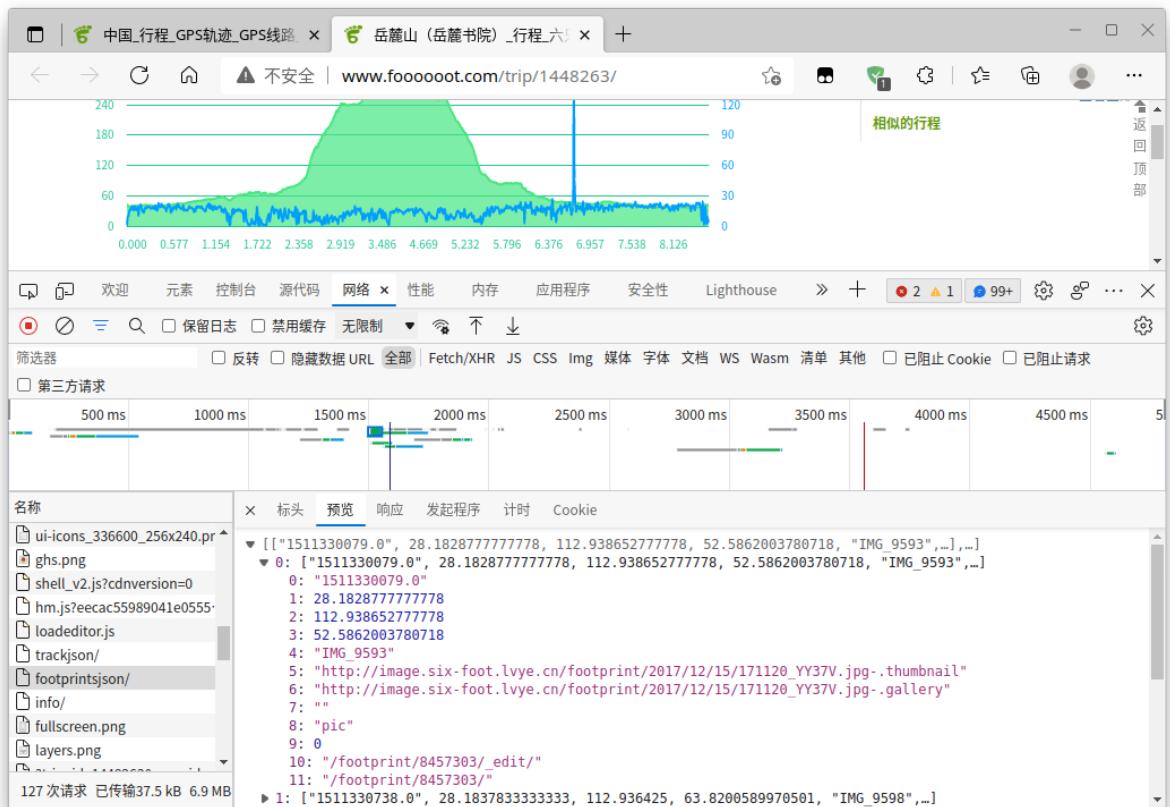
对于trackjson，前三列个数据项可以快速判断为时间戳和经纬度，对于后面三个数据项，结合网页数据：

The screenshot shows a browser window with two tabs open. The active tab is titled '岳麓山 (岳麓书院) \_行程\_六' and displays a green line graph representing a GPS track. A tooltip on the graph provides real-time data: 海拔(米): 41, 速度(公里/时): 9.4, 岔程(公里): 0.010. The x-axis represents distance in kilometers, ranging from 0.000 to 8.126. The y-axis represents altitude in meters, ranging from 0 to 240. To the right of the graph, there is a button labeled '相似的行程' (Similar Journeys). Below the main content, the browser's developer tools are visible, specifically the Network tab which shows various requests and their latencies.

可以判断分别为高程，速度和里程

对于footprintsjson，可以判断前几列数据项分别为时间戳、经纬度、高程、名字、缩略图、详情

图，后面几项笔者认为没啥作用



### 3. 数据爬取

经过上面的数据分析，爬取轨迹数据主要就是通过page和keyword构造网址获取轨迹ID，通过轨迹ID构造地址获取trackjson和footprintsjson

笔者此处基于Python，使用requests库发送http请求，使用Xpath解析界面提取数据

#### 3.1 引入库

```
In [1]:  
import requests  
from lxml import etree  
import json  
import time
```

- 注意：如果缺少相关库，请使用pip或者conda安装

#### 3.2 获取所有轨迹ID

```
In [2]:  
page_num = 1  
track_num_arr = []  
keyword = "岳麓山"  
headers = {'user-agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/62.0.3160.107 Safari/537.36'}  
page_url = "http://www.fooooooot.com/search/trip/all/1/all/time/descent/?page=" + str(page_num)
```

- 注意：发起请求请务必设置休眠时间，瞬间的大量访问请求会导致网站崩溃
- 目前此网站设置了简易反爬措施，不设置'user-agent'将无法访问

```
In [9]:
```

```

next_page = True
while(next_page):
    response = requests.get(page_url,timeout=5, headers=headers)
    tree = etree.HTML(response.text)
    trip_list = tree.xpath('//p[@class="trip-title"]/a/@href')
    if(len(trip_list) == 30):
        page_num = page_num + 1
        page_url = "http://www.fooooooot.com/search/trip/all/1/all/time/descent/?page=" + str
    else:
        next_page = False
    for trip in trip_list:
        track_num_arr.append(trip.split('/')[-2])
    time.sleep(6)

```

In [10]: `print(len(track_num_arr))`

1170

可以看到有1170条轨迹数据

### 3.3 获取轨迹数据

In [37]: `num = 0`  
`for track_num in track_num_arr:`  
 `try:`

```

#设置重连次数
requests.adapters.DEFAULT_RETRIES = 5
s = requests.session()
# 设置连接活跃状态为False
s.keep_alive = False
time.sleep(6)
footprint_url = "http://www.fooooooot.com/trip/" + str(track_num) + "/footprintsjson/"
trackjson_url = "http://www.fooooooot.com/trip/" + str(track_num) + "/trackjson/"
footprint_res = requests.get(footprint_url,headers=headers,stream=False,timeout= 10)
trackjson_res = requests.get(trackjson_url,headers=headers,stream=False,timeout= 10)
try:
    trackjson = json.loads(trackjson_res.text)
    footprint = json.loads(footprint_res.text)
    with open("./trackdata/origin/trackjson" + str(track_num) + ".json","w") as tf:
        json.dump(trackjson,tf)
    with open("./trackdata/origin/footprint" + str(track_num) + ".json","w") as ff:
        json.dump(footprint,ff)
    for track in trackjson:
        with open("./trackdata/trip_" + str(track_num) + ".txt","a") as tf:
            tf.write(str(track[1]) + " " + str(track[2]) + " " + str(track[3]) + " "
        with open("./trackdata/all.csv","a") as af:
            af.write(str(num) + "," + str(track[2]) + "," + str(track[1]) + "," + str(track[3]) + "\n")
            num = num + 1
    # print("DONE: " + track_num)
    # 关闭请求 释放内存
    footprint_res.close()
    trackjson_res.close()
    del(footprint_res)
    del(trackjson_res)
except Exception as we:
    print(we)
    print("ERROR: " + track_num)
    with open("./trackdata/error.txt","a") as af:
        af.write(str(track_num) + '\n')
    # 关闭请求 释放内存
    footprint_res.close()

```

```
trackjson_res.close()
del(footprint_res)
del(trackjson_res)
except Exception as ce:
    print(ce)
    time.sleep(60)
```

```
Expecting value: line 1 column 1 (char 0)
ERROR: 3541376
Expecting value: line 1 column 1 (char 0)
ERROR: 3541373
Expecting value: line 1 column 1 (char 0)
ERROR: 3541372
Expecting value: line 1 column 1 (char 0)
ERROR: 3541371
Expecting value: line 1 column 1 (char 0)
ERROR: 3541430
Expecting value: line 1 column 1 (char 0)
ERROR: 3505289
Expecting value: line 1 column 1 (char 0)
ERROR: 5135959
Expecting value: line 1 column 1 (char 0)
ERROR: 3390423
Expecting value: line 1 column 1 (char 0)
ERROR: 3389498
Expecting value: line 1 column 1 (char 0)
ERROR: 3392149
Expecting value: line 1 column 1 (char 0)
ERROR: 3392065
Expecting value: line 1 column 1 (char 0)
ERROR: 3392040
```

- 笔者遇到过IO错误和连接错误，所以用try except包裹起来，但是两层try总觉得不对劲
- trip\_track\_num.txt 这个文件写法主要是参考map construction的轨迹文件
- 经笔者验证， ERROR 的那几个轨迹确实没有数据

## 4. 可视化

在QGIS中利用加载XY文件的方式加载all.csv文件，并设置OSM底图，预览GPS轨迹：

