

中国各省市疫情数据的获取 (Python)

工程文件夹

DailyCrawl

工程文件目录

PSDD (D:) > 2022 > Embrace2022 > CrawlerSimple > start >

名称	修改日期
__pycache__	2022/3/20 10:47
crawled_raw_data	2022/3/20 8:53
disposed_data	2022/3/20 8:53
logs	2022/3/20 8:53
referdoc	2022/3/20 8:53
data_wash.py	2022/4/22 16:11
datastore.py	2022/3/20 10:46
main.py	2022/1/16 15:02
parameters.py	2022/1/16 15:13
wy163_crawler.py	2022/4/23 16:20

主要代码清单 215行

1. wy163_crawler.py ----- 网页抓取程序 64行
2. parameters.py ----- 参数配置 42行
3. data_wash.py ----- 数据清洗整理 61行
4. datastore.py ----- 数据入库 38行
5. main.py ----- 主控制程序 10行

其它文件

1. 爬取的原始网页内容
2. 过滤规整化的数据
3. 程序运行日志

新冠疫情数据可视化系统 (Python HTML JS)

工程文件夹

CovViewer

Python脚本清单 150行

1. app.py ----- flask应用主程序 54行
2. tools.py ----- 数据存储转化 96行

其它文件

1. HTML文件 --- 手写 $46+93+172+49=360$
2. 图片、JS文件

工程文件目录

PSSD (D:) > 2022 > Embrace2022 > CrawlerSimple > BV177411j7qJ > CovViewer >

名称	修改日期	类型	大小
^			
.idea	2022/3/4 10:04	文件夹	
__pycache__	2022/3/4 10:04	文件夹	
static	2022/3/4 10:04	文件夹	
templates	2022/3/4 10:04	文件夹	
app.py	2022/5/4 15:12	Python File	
run.cmd	2022/5/7 11:35	Windows 命令脚本	
tools.py	2022/3/18 21:21	Python File	

深圳市患者详细数据

工程文件夹

SZCrawl

工程文件目录

PSSD (D:) > 2021HaJiMeMaXiDe > Bye2021HITSZ > 2022-1

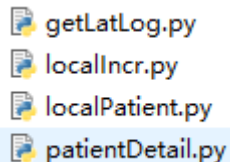
名称	修改日期
.idea	2022/3/29 14:45
__pycache__	2022/4/1 10:38
dataWashed	2022/4/1 15:19
passage	2022/4/1 9:47
sz	2022/4/1 9:58
wash	2022/4/1 15:17
Crawler.py	2022/4/2 8:58
szRun.py	2022/4/2 9:27
tmp.py	2022/4/2 9:22

其它文件

爬取过程中产生的数据文件以及
预处理过程中的中间文件等

Python脚本清单 约300行

1. Crawler.py ----- 可定制的爬虫类 114行
2. szRun.py ----- 数据爬取与预处理流程说明 32行
3. 预处理部分 ./wash 45 + 23 + 28 + 61 = 157行



getLatLog.py
localIncr.py
localPatient.py
patientDetail.py

SEIR模型与LSTM模型训练 (Python)

工程文件夹

Covid19

工程文件目录

PSSD (D:) > 2021HaJiMeMaXiDe > Bye2021HITSZ > 2022-Mr

名称	修改日期
.idea	2022/5/17 15:32
data	2022/4/28 13:29
SEIR	2022/4/28 13:29
test	2022/5/2 10:58
Visualization	2022/4/28 13:30

程序代码量 约670行

工程文件说明

1. ./data ----- 数据文件夹，存放了深圳市疫情数据
2. ./SEIR ----- 模型仿真与训练文件夹 561
SEIR模拟: $57 + 29 + 51 = 137$
SEIR参数训练: $39 + 35 + 83 + 83 = 240$
LSTM模型训练: $16 + 24 + 20 + 124 = 184$
3. ./Visualization ----- 可视化分析 110

Hawkes时空点过程模型训练 (Python)

数据预处理部分

dataproc

工程文件目录

› PSSD (D:) › 2021HaJiMeMaXiDe › Bye2021HITSZ › 2022-Mr

名称	修改日期
.idea	2022/4/9 11:53
logs	2022/4/9 12:34
rawdata	2022/4/9 11:52
referdata	2022/4/15 9:07
china_cases.npz	2022/4/18 10:55
china_cov_data.csv	2022/4/17 22:14
city_area.py	2022/4/17 21:05
cov_data_pre_process.py	2022/4/16 18:06
merge_data.py	2022/4/17 22:14
process_data.py	2022/4/18 10:55

程序代码量 约390行

工程文件说明

主要包含了中国行政区划信息爬取清洗程序以及中国各省市疫情数据过滤整合程序

代码量: $138+42+57+153 = 390$

Hawkes时空点过程模型训练 (Python)

程序代码量 约200行

主工程文件夹

neural_stpp

工程文件目录

.idea	2022/4/6 8:38
__pycache__	2022/4/8 11:02
assets	2022/4/6 8:02
build	2022/4/6 16:44
data	2022/4/6 8:02
diffeq_layers	2022/4/6 8:02
experiments	2022/4/8 11:05
flow_layers	2022/4/6 8:02
mei	2022/4/7 13:33
models	2022/4/6 8:02
CODE_OF_CONDUCT.md	2022/4/6 8:02
CONTRIBUTING.md	2022/4/6 8:02
data_utils_fast.c	2022/4/6 16:44
data_utils_fast.cp37-win_amd64.pyd	2022/4/6 16:45
data_utils_fast.pyx	2022/4/6 8:02
datasets.py	2022/4/6 8:02
iterators.py	2022/4/6 8:02
LICENSE	2022/4/6 8:02
MHP.py	2022/4/6 8:02
README.md	2022/4/6 8:02

工程文件说明

这是FaceBook基于时空点过程模型对美国新泽西州疫情数据、美国纽约市自行车行程数据以及美国地震数据进行实验分析的项目仓库，论文地址

(<https://arxiv.org/abs/2011.04583>)

本文基于中国的疫情数据使用了（神经）Hawkes模型结合HITSZ硬件资源平台进行了训练，附加代码量：200行