

团队名称：

An okay team

运行环境：

Ubuntu 18.04.3 LTS

python 3.6.8

tensorflow 2.0.0

tensorflow.keras 2.2.4-tf

numpy 1.19.0

代码说明：

pre\_processing.ipynb, 数据预处理

training.ipynb, 数据读取, 模型训练及预测

python/transform.py, 对数据进行变换时所需的函数

python/models.py, 训练时需要读取的模型, 此次只用到了 model\_TCN

python/tcn.py, models.py 里 model\_Keras\_TCN 所用的模型, 也可通过 pip 安装

运行步骤：

1. 将数据解压至文件夹 dataset/train\_data\_all
2. 运行 pre\_processing.ipynb 中所有内容
3. 运行 training.ipynb 中所有内容
4. 输出结果将自动保存在 outputs/submission.csv

思路：

我们通过观察每日感染人数, 发现其下降阶段呈指数递减趋势, 所以决定用一维线性模型对其进行拟合及预测。

由于下降之前的阶段呈非线性趋势, 我们使用 causal dilated 卷积神经网络对此阶段进行预测。当感染人数呈下降趋势后, 代码自动切换为线性模型。

神经网络的 feature 我们选用了：

- 滞后一天的每日感染人数
- 滞后一天的累计感染人数
- 滞后一天的 (第  $n$  日累计人数/第  $n-1$  日累计人数) 的倒数
- 人口相关变量：因数据未包括人口, 使用人流量的前 6 个数据的均值代替
- 迁移指数：迁进与迁出的平均, 将城市间迁移指数按城市内网格关联强度进行分配
- 人流量：因数据不全, 用二次多项式将缺失数据进行补全

模型预测 1 天的感染人数, 随后对当日的部分 feature 进行计算。不可计算的 feature 使用其最后几日的中值。得到当日所有 feature 后, 对下一日进行预测。

参考：

- 基于飞桨 PGL 的基线系统，<https://aistudio.baidu.com/aistudio/projectdetail/464528>。取数据预处理部分并进行了调整
- keras-tcn，<https://github.com/philipperemy/keras-tcn>。tcn.py 的来源，但未完成调试。
- TimeSeries\_Seq2Seq，[https://github.com/JEddy92/TimeSeries\\_Seq2Seq](https://github.com/JEddy92/TimeSeries_Seq2Seq)。与 keras-tcn 相似，但结构简单便于调试。提交时所用模型来自 TS\_Seq2Seq\_Conv\_Full\_Exog.ipynb。