

COVID-19数据分析与预测

大家都知道现在正在全球爆发的COVID-19(之前叫2019-nCoV, 新型冠状病毒肺炎等等)。这里给主WHO统计的从1月22日到今天, 也就是3月16日的全世界患病数据, 包括了累计确诊, 累计死亡, 累计治愈数据。



具体数据说明如下：

- Sno - 编号
- ObservationDate - 观测日期 in MM/DD/YYYY
- Province/State - 省或者州, 有时候一个国家已统计就可能为空
- Country/Region - 国家
- Last Update - 最后更新时间, 格式有点乱
- Confirmed - 到当前时刻累计确诊数量
- Deaths - 到当前时刻累计死亡数量
- Recovered - 到当前时刻累计治愈数量

大家可以利用这些数据做一些简单的研究和建模工作, 这个数据特别是和大家联系数据预处理, 数据可视化和简单的回归, 大家也可以参考一些流行病学的模型, 用这个数据来fit, 是一个综合调研与数据分析的联系, 并写一篇论文。具体用这些数据做什么, 我们不给出具体的任务, 但是给出以下几条建议：

1. 完成基本的Exploratory Data Analysis, 用图形、表格、动画展示你觉得有用的各种病情数据, 并给出初步的分析。
2. 建立对未来病情的发展进行预测, 实际上你在做这个题的时候已经过了很久了, 可以用现在的数据作为测试集测试拟建的模型的性能。
3. 评估一下世界各国、国内各省病毒发展时候主要采取了哪些措施, 哪些地区采取的是类似的措施, 评估这些措施的有效性, 可以结合其他的你收集的数据作为辅助, 给出如何应对病毒的方法。
4. 可用3月以前的数据作为训练, 之后的或者最新的数据作为测试。数据每天都在更新, 可在下面的链接找到。

5. 完成一篇论文。

[数据下载](#)

本题灵感来源：<https://www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset>

这上面有很多人已经完成的分析，可以参考，请勿剽窃

联系人：郑玮 qq 330839459