

修正 SEIR 动力学模型与 Logistic 模型应用

——对 COVID-19 疫情预测与评估

目录

摘要.....	
1. 问题的介绍与背景.....	4
1.1 什么是新型冠状病毒.....	4
1.2 SARS 病毒.....	4
1.3 病毒再生繁殖数 R_0	4
2. 问题分析.....	4
3. 假设.....	4
4. 建立模型.....	4
4.1 模型一: SEIR 模型.....	4
4.1.1 数据来源与问题分析.....	4
4.1.2 模型假设.....	5
4.1.3 符号说明.....	5
4.1.4 参数赋值.....	5
4.1.5 模型图.....	5
4.1.6 建立模型.....	6
4.1.7 结果分析.....	7
4.1.8 优点与缺点.....	8
4.2 模型二: Logistic 模型预测.....	8
4.2.1 数据来源与问题分析.....	8
4.2.2 模型假设.....	8
4.2.3 符号说明.....	8
4.2.4 建立模型.....	9
4.2.5 模型结果.....	9
4.2.6 结果分析.....	11
4.2.7 优点与缺点.....	11
5. 结论分析与模型改进方案.....	11
6. 建议.....	12
7. 参考文献.....	14
8. 附录.....	14
程序与代码.....	14

摘要

回顾人类的历史，其实也是跟传染病做斗争的历史。我们通过之前的数据，预测未来疫情的走势和疫情结束的时间，并分析新型冠状病毒与 SARS 病毒的异同点，提出相关建议。

模型一：在一定的假设条件下，建立优化 SEIR 模型，通过患病者、潜伏者、易感人群、康复者、死亡者之间的关系对 SEIR 模型进行优化，最终得到患病者人数与时间的关系图，发现患病者人数先是上升趋势，后出现拐点呈现下降趋势，最后 95 天后逐渐消失，可预测将在 4 月 30 日左右感染病毒的人数下降为 0。

模型二：利用 Logica 模型预测国内外的疫情情况，预测国内最大感染人数与当前累积感染人数相差不大，疫情得以控制，但是国外疫情高峰才刚刚开始，根据不同程度的防范措施，预测国外未来的最大感染人数最小为 124480 人，最大为 1500906 人。

根据以上的预测数据，认识到新型冠状病毒肺炎的严重性，在不加以管控的前提下，患者人数会呈指数上升，我们以上模型虽然存在一定的偏差，但是有一定的可靠性。

关键字：新型冠状病毒，SEIR 模型，Logistic 回归，预测

1. 问题介绍与背景

1.1 什么是新型冠状病毒^[1]

因 2019 年武汉病毒性肺炎病例被发现，2020 年 1 月 12 日被世界卫生组织命名为 2019 新型冠状病毒（2019-nCoV），该病毒是一种急性感染性肺炎，其病原体是先前未在人类中发现的新型冠状病毒，根据感染病毒的人会出现不同程度的症状，有的病例临床表现为发热、呼吸困难、咳嗽，腰酸背痛等，令人奇怪的是，有的人虽然没有一点症状，但是通过检测显示已经被感染，体现出该病毒的不确定性。

该病毒的传播途径至今还未十分明确，可以确定的传播途径包括呼吸道飞沫传染、接触传染与气溶胶传播。经呼吸道飞沫传播，如患者打喷嚏、咳嗽、说话的飞沫，呼出的气体或者近距离接触，都有被传染的风险，患者的飞沫如果混合空气，形成气溶胶，他人吸入后也可导致感染。根据资料显示，病毒在光滑的物体表面可以存活数个小时，温度湿度合适的环境能存活一天，有发现可达到五天的病毒，这给我们抗击病毒带来障碍。

1.2 SARS 病毒^[2]

SARS 病毒是冠状病毒的一种，2002 年在我国流行的非典型肺炎的病原体，简称非典病毒。该病毒与流感病毒十分相似，但又具有独特的致病性，在 2002 年到 2003 年在我国肆虐，传染性极强，发病急，出现严重的急性呼吸窘迫综合征、呼吸衰竭，病死率非常高。

该病毒在世界范围内分布很广，能够广泛感染人和动物病毒，对温度比较敏感，不易生长。SARS 病毒是成人普通感冒的病原菌之一，主要通过呼吸道感染，很少波及下呼吸道，如果迁延不愈可以出现重症肺炎，呼吸道分泌物可以经口排出，通过空气飞沫传播，所以要做好防护，及时打流感疫苗。

1.3 病毒再生繁殖数 R_0 ^[3]

这是传染学上一个专门用来衡量一个病毒的传染性强弱的指标，指的是一个感染者在他的整个传染期通过与易感染者有效接触导致其感染的期望数。不同的病毒， R_0 值越高，说明病毒传染能力越强，疫情越严重，控制的难度越大。

- 当病毒的 $R_0 > 1$ 的时候，意味着一个感染者会传染超过一个人，病毒呈扩散态势；
- 当病毒的 $R_0 = 1$ 的时候，意味着一个感染者会传染一个人，病毒持续传播；
- 当病毒的 $R_0 < 1$ 的时候，意味着一个感染者会传染少于一个人，疫情将逐渐消退；

说明： R_0 并不是一成不变的，在传染初期，确诊人数较少，政府和群众往往不会立即引起重视，这个阶段下，疫情没有加以控制，病毒传播速度会加快，

所以病毒爆发扩散期 R_0 值会较高。等到政府和医院意识到事情的严重性时，纷纷各省出现管控措施，有效的防护隔离一旦落实，病毒传染的可能性大大降低，感染人数会有一定的减少，这时 R_0 会随之下降。当 $R_0 < 1$ 的时候，疫情也就被控制住，直至 R_0 最后为 0，疫情才得以彻底控制。

2. 问题分析

针对问题一和问题二，可以分为以下两部分：

(一) 建立相应的数学模型来预测病毒的趋势

- 根据新型冠状病毒近来的数据，运用 matlab 软件建立出相应的 SEIR 模型，用模型结果来预测国内新型冠状病毒未来的趋势；
- 运用 python 软件建立 Logistic 增长模型，通过该模型来拟合当前病毒感染的人数来预测未来国内外的感染人数的增长变化及感染极限；

(二) 比较 SARS 病毒与新型冠状病毒的异同点

根据资料的查找，找到两种病毒的异同点进行比较，从而结合实际情况，对社会和国家公共卫生管理提出合理的建议。

3. 假设

通过以上的分析，为了便于问题的研究，不妨忽略一些对问题影响较小的因素，做如下的假设：

- 模型建立的过程中不考虑感染者的迁入与迁出；
- 模型中忽略防治政策变化对参数变化的影响；
- 所得数据虽有一定误差，但是真实有效；

4. 建立模型

4.1 模型一：SEIR 模型

4.1.1 数据来源与问题分析

模型一采用的疫情数据来源于百度实时疫情报告。

新型冠状病毒是一种具有潜伏期的传染性病毒，新型冠状病毒主要通过呼吸道飞沫传播和接触传播，各个年龄段都可能被传染，为了了解疫情的未来发展，因此我们需要建立一个模型来预测病毒传染趋势。

4.1.2 模型假设

- 全国人都基本处于隔离状态;
- 潜伏者假定没症状或者症状较轻,有一部分具有传染性,患病者具有症状与传染性;
- 所有健康的人都为易感人群;
- 忽略国内与国外人口流动情况;
- 患病者恢复后会产生抗体,即康复的人不会二次感染;

4.1.3 符号说明

I : 患病者, S : 易感人群, E : 潜伏者

R : 康复者, D : 死亡者, P : 总人数

α_0 : 患病者让易感人群变为潜伏者的概率;

α_1 : 潜伏者让易感人群变为潜伏者的概率;

β_0 : 潜伏者变为患病者的概率;

β_1 : 潜伏者变为易感人群的概率;

γ_0 : 患病者死亡的概率;

γ_1 : 患病者康复的概率;

t_0 : 患病者平均每天遇到的易感人群人数;

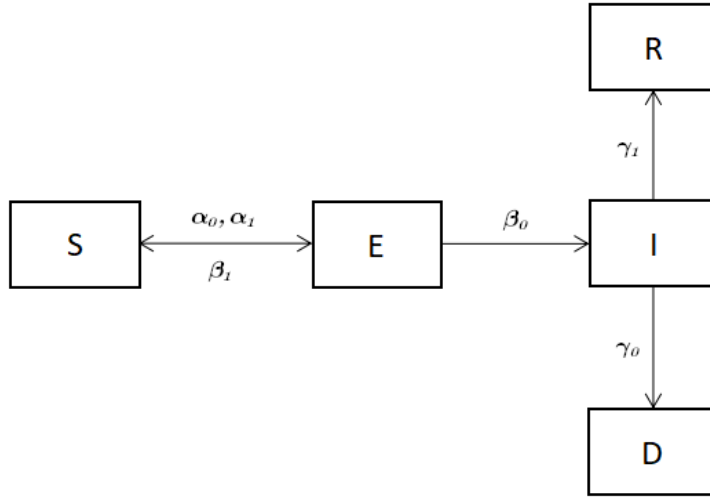
t_1 : 潜伏者平均每天遇到的易感人群人数;

4.1.4 参数赋值

我们取 1 月 26 号为开始起点,初始患病人数为 2630 人,潜伏者为 5794 人,康复者为 51 人,死亡人数为 80 人,全国总人数约为 14 亿。2 月 2 号,群众基本处于隔离状态,患者在医院中,隔离程度更高,因此将 t_0 设为 1, t_1 设为 3。

基于原始数据对参数进行多次优化,最终将 α_0 设为 0.78, α_1 设为 0.205, 将 β_0 设为 0.25, β_1 设为 0.70, γ_0 取 0.035, γ_1 开始取 0.0185, 之后每天 γ_1 增加 0.02, 到 0.8 后保持不变。通过这些参数建立出了较为符合的模型。

4.1.5 模型图



4.1.6 建立模型

通过对疫情态势的了解，以及对过往的数据的分析，为了准确预测出未来疫情的发展，我们建立出以下 SEIR 模型的方程：

$$\frac{dS}{dt} = -t_0\alpha_0IS - t_1\alpha_1ES + \beta_1E$$

$$\frac{dE}{dt} = t_0\alpha_0IS + t_1\alpha_1ES - \beta_0E - \beta_1E$$

$$\frac{dI}{dt} = \beta_0E - \gamma_0I - \gamma_1I$$

$$\frac{dD}{dt} = \gamma_0I$$

$$\frac{dR}{dt} = \gamma_1I$$

得到：

$$S(t+1) = S(t) - \frac{t_0\alpha_0I(t)S(t)}{P} - \frac{t_1\alpha_1E(t)S(t)}{P} + \beta_1E(t)$$

$$E(t+1) = E(t) + \frac{t_0\alpha_0I(t)S(t)}{P} + \frac{t_1\alpha_1E(t)S(t)}{P} - \beta_0E(t) - \beta_1E(t)$$

$$I(t+1) = I(t) + \beta_0E(t) - \gamma_0I(t) - \gamma_1I(t)$$

$$D(t+1) = D(t) + \gamma_0I(t)$$

$$R(t+1) = R(t) + \gamma_1I(t)$$

4.1.7 结果分析

运用 matlab 软件得到结果：

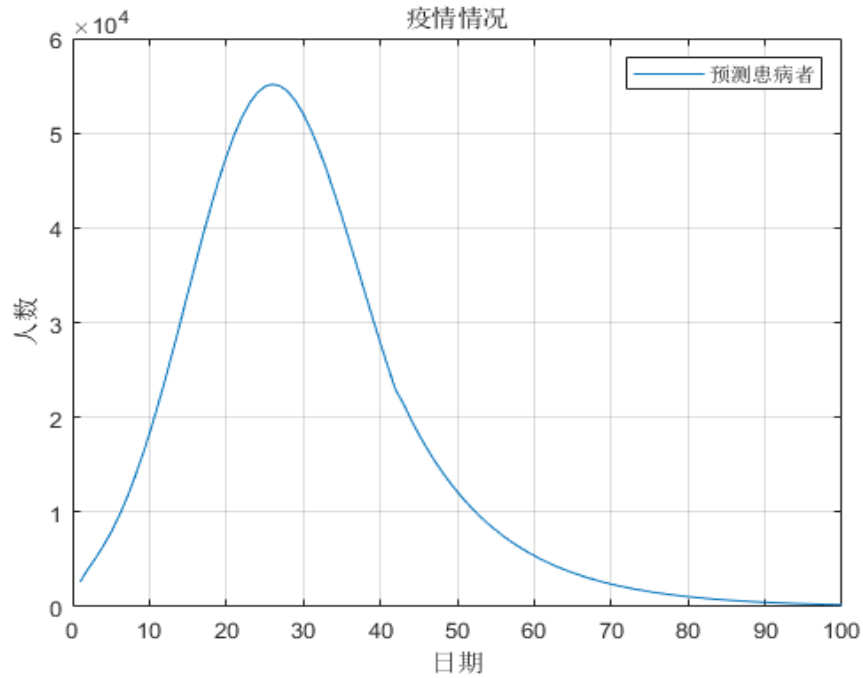


图 4-1

根据图 4-1 可预测在第 26 天（2 月 20 日）前感染病毒的人数将会以较快的速率持续增加。在 2 月 20 日左右将会出现拐点，感染病毒的人数将会持续下降，下降的速率由快到慢，最后在第 95 天（4 月 30 日）左右感染病毒的人数消失。

我们将实际数据与预测数据进行对比：

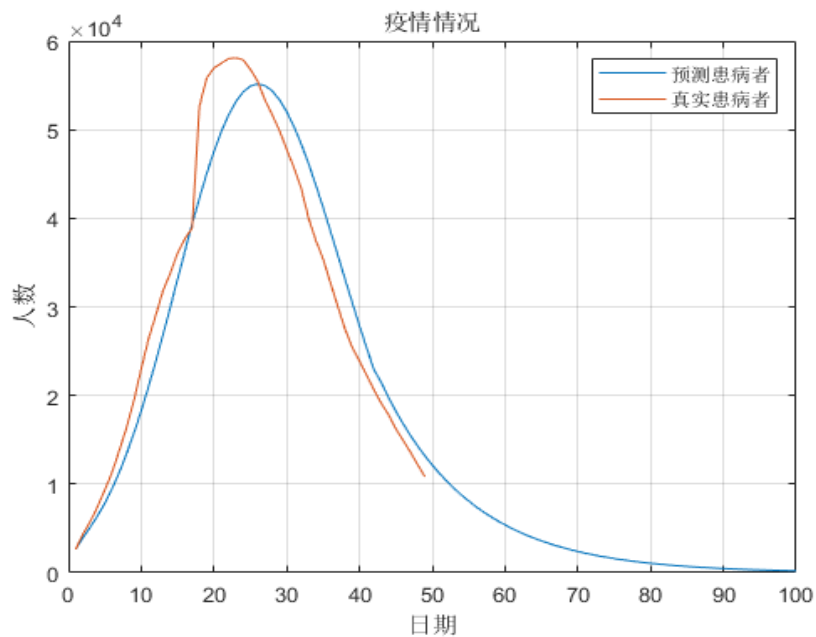


图 4-2

通过图 4-2 得到预测数据与实际数据具有相似的趋势，感染最高人数有所差别，但总体相似，可说明预测数据具有一定可信性，证明该 SEIR 模型对疫情走势的评估具有一定可靠性。

4.1.8 优点与缺点

优点：

- 与真实数据具有相似的趋势，有较好的预测能力；
- 模型具有坚实可靠的数学基础；
- 模型简单，思路清晰，易于实现；

缺点：

- 模型结构较为单一，只考虑了患病者的数据；
- 忽略了较多因素，比如隔离，政策，迁入迁出等等；
- 某些参数具有较大的主观因素；

4.2 模型二：Logistic 模型预测

4.2.1 数据来源与问题分析

新型冠状病毒作为一种全新的病毒，它感染人类的数学模型类似一个新物种迁入到一个新生态系统中后，其数量变化的数学模型。运用 Python 程序建立 Logistic 增长模型，通过该模型来拟合当前病毒感染人数来预测新型冠状病毒肺炎国内外的感染人数的增长变化及感染极限。国内数据来源于百度新型冠状病毒肺炎疫情实时大数据报告，国外数据来源于澎湃新型冠状病毒感染肺炎病例实时地图。

4.2.2 模型假设

- 国内以 1 月 26 号的病毒累积感染人数作为病毒初始感染人数；
- 国外以 2 月 15 号的病毒累积感染人数作为病毒初始感染人数；

4.2.3 符合说明

K 为环境容纳量，是指在自然条件下（环境资源总是有限的），在有限的环境资源条件下，能够达到的最大种群密度。在本模型中指新型冠状病毒累积最大的感染人数。

P_0 为初始容量，就是 $t=0$ 时刻的数量，即开始时刻的病毒感染人数。

r 为增长速率， r 越大则增长越快，越快逼近 K 值， r 越小增长越慢，越慢逼近 K 值。

4.2.4 建立模型

Logistic 方程：

$$P_{(t)} = \frac{KP_0 e^{rt}}{K + P_0(e^{rt} - 1)}$$

4.2.5 模型结果

根据以上的模型，用 python 写代码得出模型的结果。

国内：

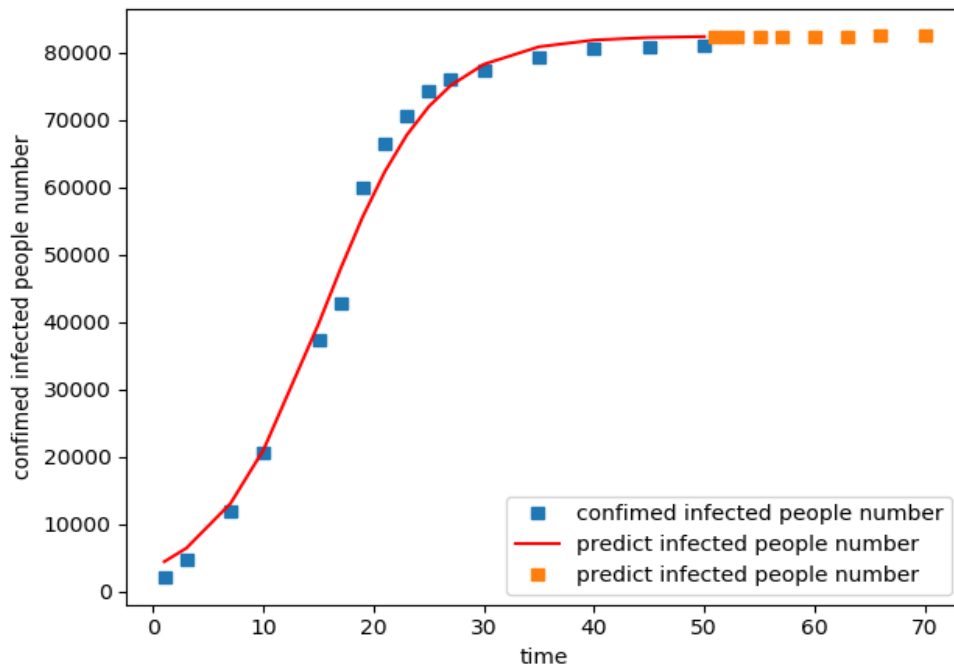


图 4-3 国内 Logistic 模型 (r = 0.2)

当 r=0.2 时，模型的拟合情况与现时情况相接近。通过 python 程序得到 K=82489，即未来最大感染人数为 82489 人。

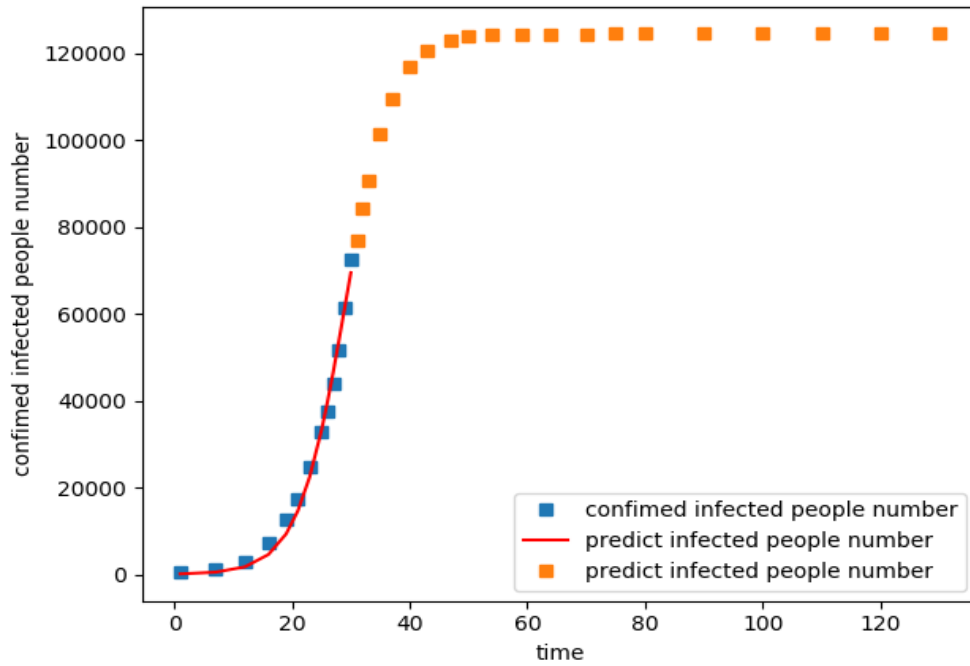


图 4-4 国外 Logistic 模型 ($r = 0.25$)

当 $r=0.25$ 时,通过 python 程序得到 $K=124480$,即未来最大感染人数为 124480 人。由图像可知当 $t \approx 60$ 时 (未来 30 天后), 达到最大感染人数。

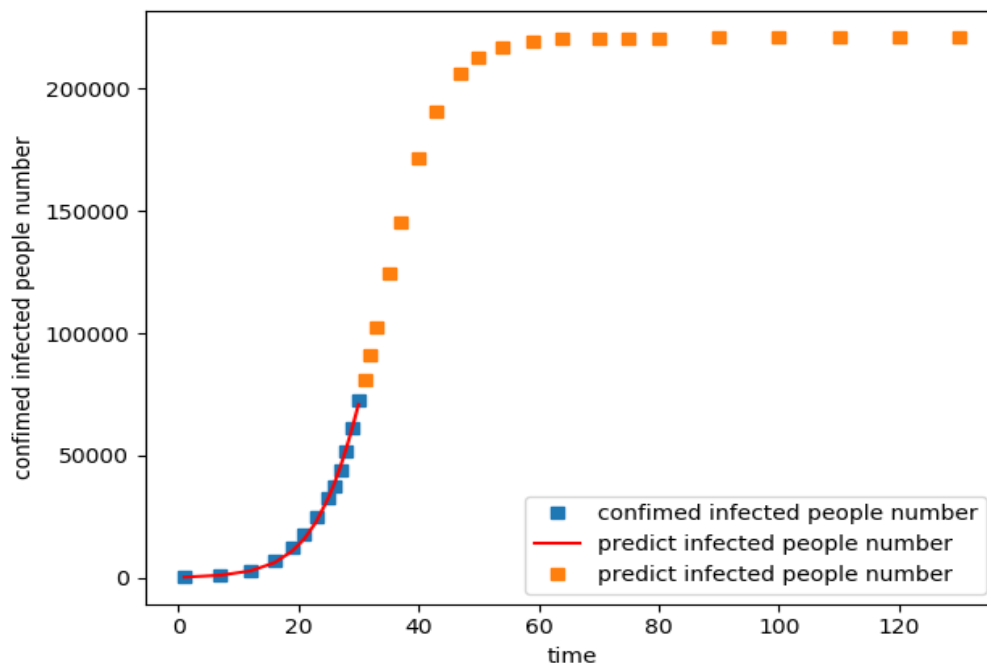


图 4-5 国外 Logistic 模型 ($r = 0.2$)

当 $r=0.2$ 时,通过 python 程序得到 $K=220809$,即未来最大感染人数为 220809 人。由图像可知当 $t \approx 80$ 时 (未来 50 天后), 达到最大感染人数。

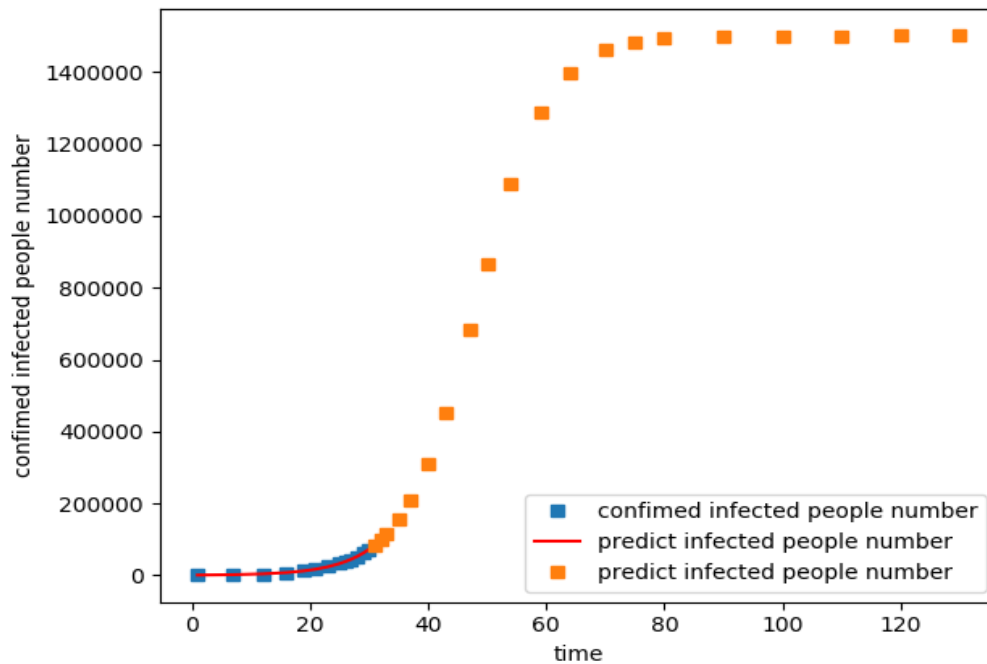


图 4-6 国外 Logistic 模型 ($r=0.165$)

当 $r=0.165$ 时，通过 python 程序得到 $K=1500906$ ，即未来最大感染人数为 1500906 人。由图像可知当 $t \approx 100$ 时（未来 70 天后），达到最大感染人数。

4.2.6 结果分析

由图像得知当前国内疫情高峰已过，且预测国内最大感染人数与当前累积感染人数相差不大。国外疫情高峰才刚刚开始，预测根据不同的防范措施，未来的最大感染人数最小为 124480 人，最大为 1500906 人。

4.2.7 优点与缺点

优点：

- 该模型较为直观，易于观察得出结论
- 通过 Python 程序计算可以直接得到未来最大感染人数

缺点：

- 该模型为预测模型，可能与未来的实际情况有出入

5. 结论分析与模型改进方案

1. 分析

模型一：在一定的假设条件下，建立优化 SEIR 模型，通过患病者、潜伏者、易感人群、康复者、死亡者之间的关系对 SEIR 模型进行优化，最终得到患病者人数与时间的关系图，发现患病者人数先是上升趋势，后出现拐点呈现下降趋势，最后 95 天后逐渐消失，可预测将在 4 月 30 日左右感染病毒的人数下降为 0。

模型二：利用 Logistic 回归预测国内外的疫情情况，预测国内最大感染人数与当前累积感染人数相差不大，国外疫情高峰才刚刚开始，根据不同程度的防范措施预测未来的最大感染人数最小为 124480 人，最大为 1500906 人。

2. 模型改进方案

由于能力有限，以上的预测存在一定的偏差，外界因素影响较大，模型一考虑的因素不够完整，还应该加入隔离潜伏者、未被发现的潜伏者与感染者、隔离易感者等类型的群体，我们利用有限的的数据，分析的结果也符合疫情的走向，结果具有一定的价值。

6. 建议

新型冠状病毒肺炎已经成为全球关注的热点。起初只在中国出现了病例，在短短不到一个月的时间里，世界各处都出现了该病症的患者，该情况与 2003 年中国的非典疫情有相似之处，新型冠状病毒肺炎与非典相比十分有必要。

经过资料的查找，我们整理成下表：

表 1 SARS 与新型冠状病毒的三个比率对比

	病死率 ^[4]	传播率 (R0) ^[4]	治愈率
SARS 病毒	9.2%	2.6	76.4%
新型冠状病毒肺炎	2.8%	3.7~4.3 (2020 年 2 月评估值)	74%

说明： 新型冠状病毒肺炎的数据由于该疫情未结束，数据仍然在不断的变化，准确数字无法得到，以上数据根据相关资料选取，为估计值。传播率我们使用 R0 大小来体现。

从上面表 1，我们可以看出新型冠状病毒比 SARS 的病死率低和治愈率相对较高，这也体现出我国医疗水平的提高和政府对于控制防护措施实施的努力，但是新型冠状病毒的传播能力大，这需要我们继续加强对疫情的防控工作，不可以轻易懈怠。

除以上的不同，我们还对病毒各个方面进行了比较：

表 2 SARS 病毒与新型冠状病毒其他方面的异同

	潜伏期(天)	易感人群	感染症状	传播途径	相同
SARS 病毒 ^[2]	2~14 天	20~60 岁之间占比最多	高热为首发症状, 呼吸衰竭、休克和多脏器功能衰竭等, 一般出现症状之后才有传染性	近距离呼吸道飞沫及密切接触传播, 还有粪口传播、尿液等不可排除。存在超级传播者	受体蛋白相同, 基因序列相似度
新型冠状病毒肺炎 ^[1]	一般 1~3 天, 有达到 14 天, 最长为 24 天	各个年龄段的人都有可能被感染儿童和孕产妇是易感人群	发烧或轻微咳嗽等, 无症状的感染者也具有传染性	呼吸道飞沫传播、接触传播, 气溶胶, 粪口传播尚待研究, 未查出是否有超级传播者	达到 80%等

从上面的表 2, 我们可以看出新型冠状病毒比 SARS”狡猾”, 新型冠状病毒潜伏期长, 易感人群广泛, 无症状感染者也会传染给其他人, 这无疑给疫情的防控工作带来难点。

根据以上资料与建立模型的预测情况, 虽然我国已经过了高峰阶段, 但是疫情仍然在进行, 并且还需持续一段时间, 我们不可懈怠。我们结合实际情况对社会和国家在公共卫生管理方面上给出以下建议^[4]:

传染源的控制

对病人要早发现、早隔离、早治疗, 对于密切接触者均实施医学观察, 隔离期结束后, 对无症状和体征的隔离观察者要及时解除隔离, 如果隔离观察者发展为新型冠状病毒, 要严格按病人实施管理, 再次对密切接触者进行追踪。

切断传染途径

在公共场所, 疫情较严重的地方要定期消毒; 对于医院, 选择符合条件的医院和病房收治病人是避免医院内感染的重要前提, 建立、健全医院内感染管理组织, 制定相关管理制度, 严格消毒, 做好医护人员的防护措施。对于群众, 呼吁大家做好个人防护, 出门戴口罩, 远离人群, 不走亲戚不串门, 回家之后要勤洗手等。

对易感人群的保护

目前疫苗还没有研制出来，对易感人群的保护也是至关重要的。政府可暂停一部分的场外活动，学校、单位进行停课停工，减少人出门的次数，也可大大降低感染的几率。

对国-国流动人口的检测

根据我们的模型可得知未来国外患病人数将急剧上升，至今为止，没有一个国家可以达到中国对疫情管控的水平，因此，建议对入境者加以调查和检查，主动隔离 14 天，确定无携带病毒之后，才允许入境，该措施使国内疫情不会大幅度的变化，疫情也会更早的彻底控制。

以上建议望采纳。

7. 参考文献

- [1] <https://baike.baidu.com/item/2019%E6%96%B0%E5%9E%8B%E5%86%A0%E7%8A%B6%E7%97%85%E6%AF%92/24267858?fromtitle=%E6%96%B0%E5%9E%8B%E5%86%A0%E7%8A%B6%E7%97%85%E6%AF%92&fromid=7904360&fr=aladdin>
- [2] <https://baike.baidu.com/item/SARS%E7%97%85%E6%AF%92/4005860?fr=aladdin>
- [3] <https://zhuanlan.zhihu.com/p/111666596>
- [4] SARS . 中国疾病预防控制中心. 2017. 12. 28

8. 附录

SEIR 模型代码:

```
real_data=[2630
4369
5762
7440
9336
11319
13799
16402
19414
22980
26343
29032
31825
33788
36043
37693
38874
```

52599

55823

56951

57493

58010

58097

57886

56810

55475

53371

51689

49910

47760

45697

43352

40011

37502

35420

32741

30096

27524

25441

23872

22263

20616

19097

17802

16226

14920

13607

12178

10822

]

P=1400000000;

I=2630;%患病人数

R=51;%治愈人数

D=80;%死亡人数

E=5794;%潜伏者

S=P-I-E-D-R;%易感人数

t0=1;

t1=3;

alpha0=0.78;

alpha1=0.205;

beta0=0.25;

```

beta1=0.70;
y0=0.035;
y1=0.0185;%康复概率
T=1:100;
for idx =1:length(T)-1

S(idx+1)=S(idx)-t0*alpha0*I(idx)*S(idx)/P-t1*alpha1*E(idx)*S(idx)/P+beta1*E(idx);%易感人数迭代

E(idx+1)=E(idx)+t0*alpha0*S(idx)*I(idx)/P+t1*alpha1*E(idx)*S(idx)/P-beta0*E(idx)-beta1*E(idx)%潜伏者人数迭代
    I(idx+1)=I(idx)+beta0*E(idx)-y0*I(idx)-y1*I(idx);%患病人数迭代
    R(idx+1)=R(idx)+y1*I(idx);%康复人数迭代
    D(idx+1)=D(idx)+y0*I(idx);%死亡患者人数迭代
    if y1<0.8
        y1 = 0.02+y1;
    else
        y1 = 0.8;
    end
end
figure(1)
plot(T,I);
grid on;
xlabel('日期');
ylabel('人数');
legend('预测患病者');
title('疫情情况');
hold on ;
plot(real_data);
xlabel('日期');
ylabel('人数');
legend('预测患病者','真实患病者');
title('疫情情况');

```

Python 代码

```

import numpy as np

import matplotlib.pyplot as plt

import math

import pandas as pd

import numpy as np

```



```
import matplotlib.pyplot as plt
from scipy.optimize import curve_fit

def logistic_increase_function(t, K, P0, r):
    t0 = 1
    r = 0.2
    exp_value = np.exp(r * (t - t0))
    return (K * exp_value * P0) / (K + (exp_value - 1) * P0)

t = [1, 3, 7, 10, 15, 17, 19, 21, 23, 25, 27, 30, 35, 40, 45, 50]
t = np.array(t)

P = [2048, 4614, 11885, 20513, 37293, 42754, 59907, 66581, 70644, 74279, 75997,
77269, 79394, 80580, 80933, 81062]
P = np.array(P)

popt, pcov = curve_fit(logistic_increase_function, t, P)
print("K:capacity  P0:initial_value  r:increase_rate  t:time")
print(popt)
P_predict = logistic_increase_function(t, popt[0], popt[1], popt[2])
future = [1, 3, 7, 10, 15, 17, 19, 21, 23, 25, 27, 30, 35, 40, 45, 50, 51, 53, 55, 60, 65,
71, 81, 91, 101]
future = np.array(future)
future_predict = logistic_increase_function(future, popt[0], popt[1], popt[2])
tomorrow = [51, 52, 53, 55, 57, 60, 63, 66, 70]
tomorrow = np.array(tomorrow)
tomorrow_predict = logistic_increase_function(tomorrow, popt[0], popt[1], popt[2])

plot1 = plt.plot(t, P, 's', label="confimed infected people number")
plot2 = plt.plot(t, P_predict, 'r', label='predict infected people number')
plot3 = plt.plot(tomorrow, tomorrow_predict, 's', label='predict infected people
number')
plt.xlabel('time')
```

```
plt.ylabel('confimed infected people number')
```

```
plt.legend(loc=0)
```

```
print(logistic_increase_function(np.array(28), popt[0], popt[1], popt[2]))
```

```
print(logistic_increase_function(np.array(29), popt[0], popt[1], popt[2]))
```

```
plt.show()
```