

基于SIR模型的新型冠状病毒肺炎预测分析

康观龙 柳炳祥

(景德镇陶瓷大学,江西 景德镇 333403)

摘要: 本文利用Python爬虫技术,获取国家卫健委每日发布的新冠肺炎相关病例(确诊、疑似、治愈、死亡)数据,基于疫情初期数据用SIR模型建立疫情变化趋势图,再结合政府制定的一系列干预政策,考虑其他影响疫情发展的因素,用改进的SIR模型对疫情进行预测,并对疫情的发展阶段和趋势进行评估,得出在政府干预下,新冠肺炎疫情高峰下降及结束时间提前,肯定了政府干预措施的必要性,为世界各国建立有效的疫情防控提供参考。

关键词: Python爬虫技术; SIR模型; 新冠肺炎疫情

中图分类号: R563.1

文献标识码: A

1 引言

2019年新型冠状病毒肺炎在全国迅速蔓延,极大影响了社会的正常运转,在国家的强力有效干预下,改善卫生设施,提高针对性的医疗技术和水平,新冠肺炎在我国已经得到有效的控制。新冠肺炎爆发以来,作者密切关注疫情发展并试图利用已有数据对疫情进行预测分析,现有SIR传染病预测模型对疫情进行预测,并没有考虑外部因素的影响,文章结合国家相关干预手段,利用改进的SEIR模型对已有疫情数据进行预测分析和评估,对比政府干预前后的疫情数据,认识到政府干预对公共安全事件的积极影响,为今后预防、控制和治疗类似疫情提供参考。

2 SIR模型建立

Kermack等利用参数优化及非线性动力学方法,建立了研究传染病传播的经典仓室模型^[1]。仓室模型的思想,就是把自然状态下的人群分为若干仓室,处于传染病不同阶段的人群代表相应的仓室。仓室模型最具代表性的模型是SIR模型, SIR模型中的S、I和R分别代表Susceptible(易感者)、Infectious(感病者)及Recovered(康复者),易感者就是指存在传染病的地区中健康但是会被该传染病感染的人群,感病者就是指在该地区因为某些原因不幸得到传染病且可能将传染病传染给他人的人群,康复者是指该地区中感病者经过治疗或者其他原因痊愈的人。

SIR模型提出的前提是:发生传染病的地区在该疾病出现至基本结束期间,人口数量没有变化即出生率和死亡率为零,人口也没有迁入迁出。所以总人口 $N=S+I+R$,另外染病者痊愈后不再被感染。由于传染病的传染性,各仓室人群的数量是不断变化的,所以用 $S(t)$ 、 $I(t)$ 、 $R(t)$ 分别表示在 t 时刻各阶段人群所处仓室的数量并建立随时间变化的SIR模型方程组:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{S}{N} I \\ \frac{dI}{dt} &= \beta \frac{S}{N} I - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}\quad (1)$$

其中 $\frac{dS}{dt}$ 是易感者 S 关于时间 t 的微分,表示单位时间(天)内易感者的变化。同理, $\frac{dI}{dt}$ 和 $\frac{dR}{dt}$ 表示的就是单位时间(天)

内感病者和康复者的变化。 β (感染率)表示的是易感者接触感病者后感染该传染病的概率, γ (治愈率)表示的是感病者经过一段时间后痊愈的概率。

求肺炎感染率时,可以用国家卫健委发布的疫情实时数据反向推理拟合,提高感染率的准确性,这是借鉴马知恩等在SARS疫情传播趋势预测中的方法^[2]。主要是利用感病者通过接触或者其他传播途径将疾病传染给易感者,易感者经过时间 T 出现感病者症状,所以人群中的感病者是 $t-1$ 到 $t-T$ 时间的新感染者, $F(t)$ 则表示在 T 时新增确诊病例,其中 T 表示该传染病的潜伏期,则有:

$$\beta(t) = \frac{F(t)}{I(t)} = \frac{\hat{F}(t+T)}{\sum_{j=0}^{T-1} (t+j)} \quad (2)$$

前面已经用Python获取了国家卫健委每日发布的新冠肺炎相关数据,很容易就能求得新冠的治愈率 β ,现在只要计算出新冠肺炎的潜伏期,就可以求出传染率,带入公式1就可以对新冠肺炎的发展进行预测。

钟南山院士团队采用1月29日前31个省市552医院的1099名患者的资料,描绘了该病的临床特征,并通过MuLBSTA评分方法,评估了病毒性肺炎患者的90天死亡风险,其结果表明新冠肺炎潜伏期的中位数为4天(当天感染当天发病的潜伏期算1天),死亡率较低^[3]。邱明悦等采用从1月13日~2月18日,共543个新冠肺炎的确诊病例。利用Bayes估计和极大似然估计两种参数估计方法在双重区间删失的数据结构下,对潜伏期分布进行统计建模,计算得出新型冠状病毒肺炎(COVID-19)的潜伏期中位数约为5.5天^[4]。综合考虑,本文取新冠肺炎的潜伏期为4.5天,得到传染率:

$$\beta(t) = \frac{F(t)}{I(t)} = \frac{\hat{F}(t+4.5)}{\sum_{j=0}^{3.5} (t+j)} \quad (3)$$

求治愈率前先解释一个概念:病程,就是疾病从发生,发展到结局的全部过程,那么 $\gamma=1/G$ 就可以认为是死亡或者治愈的平均速率,所以治愈率取决于感染的平均时间 G ,即医学意义上的总病程。

基于武昌方舱医院364位患者恢复数据得到的结果^[5],总病程选择 $G=28.1d$,则 $\gamma=1/G=1/28.1$ 。具体总病程分布情况如

收稿日期: 2020-06-04

作者简介: 康观龙(1996-),男,硕士研究生,研究方向:大数据分析;柳炳祥(1966-),男,教授,硕士生导师,研究方向:数据挖掘与统计分析。

图1所示。

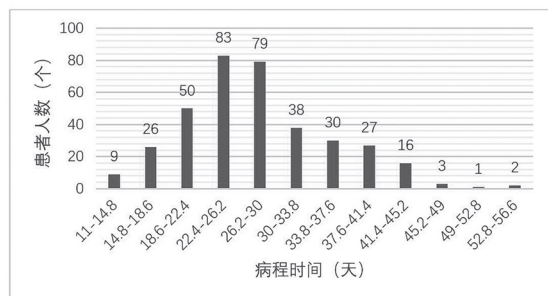


图1 武昌方舱医院患者总病程分布情况

将得到的传染率和治愈率带入公式中，得出疫情趋势图如图2所示。

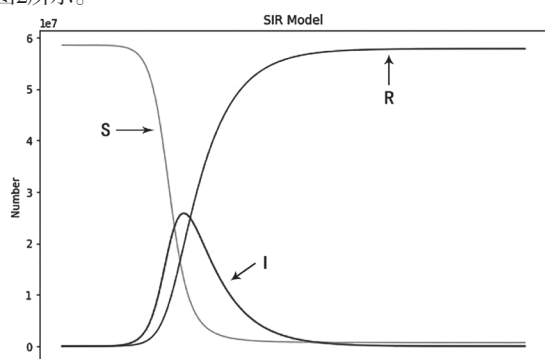


图2 疫情趋势

由图2可以得出，在不存在外部因素的干扰下，新冠肺炎爆发后，感病者开始增加，大约在30天是发病高峰，之后逐渐减少，80天后疫情结束。比较根据国家卫健委发布的病例实时数据和累积数据画的趋势图（图3，图4），发现SIR模型得出的趋势判断较实际数据有很大出入。

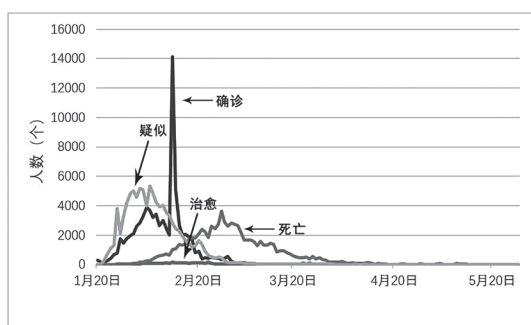


图3 全国当日病例

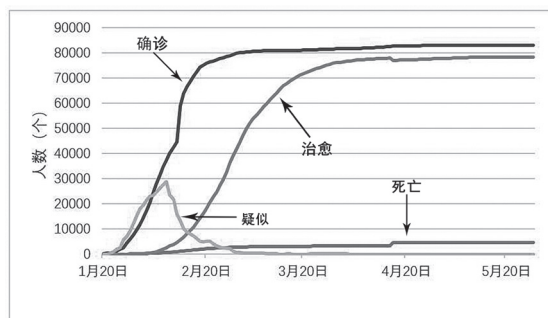


图4 全国累计病例

3 SIR模型改进

传染病周期较长，染病者因病死亡、政府政策干预、注射疫苗等因素都会对传染病的发展产生极大的影响，单纯的SIR模型存在一定的局限性。所以文中提出SEIR模型，模型中新增加了潜伏者，同时感病者和潜伏者都因为具有传染性而对易感者产生威胁。SEIR模型就是在SIR的模型上加入E（潜伏者）仓室，由于新冠肺炎的传播存在潜伏期，易感者接触感病者后并不会立即出现发病症状，而是经过一段时间的潜伏再发病变成感病者，结合SIR模型，对每一个仓室的人群随时间变化建立平衡方程式，得到如下SEIR模型方程组：

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{S}{N} E - \epsilon \frac{S}{N} I \\ \frac{dE}{dt} &= \beta \frac{S}{N} E + \epsilon \frac{S}{N} I - \delta E \\ \frac{dI}{dt} &= \delta E - \gamma I \\ \frac{dR}{dt} &= \gamma I \end{aligned} \quad (4)$$

其中 $\frac{dS}{dt}$ 、 $\frac{dE}{dt}$ 、 $\frac{dI}{dt}$ 和 $\frac{dR}{dt}$ 表示易感者、潜伏者、感病者和恢复者在一段时间（天）内的数量变化。式中 β 表示该段时间内潜伏者接触易感者并将其感染的概率， ϵ 表示的是该段时间内感病者接触易感者并将其感染的概率（ $\epsilon > \beta$ ）， δ 表示被感染但未出现感病者症状的潜伏者在该段时间变成感病者的概率， γ 表示的是感病者经过一段时间后痊愈的概率。

由于模型参数较多，引入初始参数正向推演疫情扩散过程，并通过新冠肺炎实际数据反演动态修正和求解得到模型最终参数估计值^[6]，相关初值查阅国家卫健委每日发布的疫情实时数据，将得到的参数带入SEIR模型的方程组（4）中。利用Python中matplotlib库得出模型预测的疫情图，如图5所示。

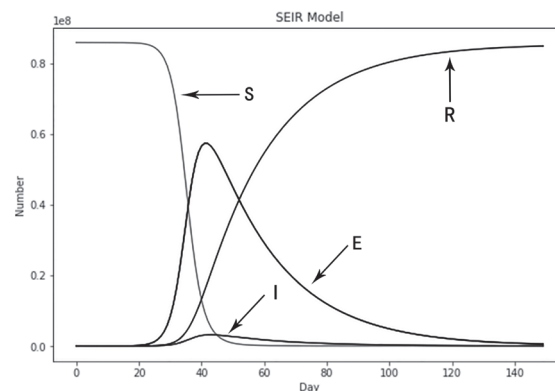


图5 疫情分布图

由图5可知，在加入潜伏者仓室后，新冠肺炎的潜伏者数量会在爆发后大约40天达到顶峰，之后逐渐消失，而感病者在潜伏者爆发后也达到最高值。SEIR模型考虑到了病毒传染过程中存在的潜伏状态，同时也考虑到了潜伏者和感病者对易感者的传染，模型中新冠肺炎的实际传染性因为两个仓室都具有传染性变强了，所以相应的潜伏者和感病者增加，完全康复时间变长。

上述SEIR模型考虑到了易感者接触感病者感染后变成潜伏者的过程，较完整的表现了一般传染病从侵入到潜伏再到发病最终康复的过程，但是我国新冠肺炎从发现到控制最后治愈，都离不开政府的干预和社会大众的理解支持，所以提出了改进的SEIR模型。

该模型增加了一个新的仓室（隔离仓室），同时考虑了

感病者因病死亡的问题,改进的SEIR模型中,感病者被发现后就被隔离了,由于我国隔离措施做的好,有理由认为潜伏者一有发病症状就被隔离了,所以感病者仓室的人群就相当于隔离者仓室,因此该传染病传播过程中易感者只被潜伏者感染,同时感病者自愈或者治愈系数代表的是感病者未隔离或已隔离痊愈的概率,另外死亡的患者也不再具备传染能力。结合SEIR模型,对每一个仓室的人群随时间变化建立平衡方程式,得到如下改进的SEIR模型方程组:

$$\begin{aligned}\frac{dS}{dt} &= -\beta \frac{S}{N} E \\ \frac{dE}{dt} &= \beta \frac{S}{N} E - \delta E \\ \frac{dI}{dt} &= \delta E - \gamma I - \mu I \\ \frac{dR}{dt} &= \gamma I \\ \frac{dD}{dt} &= \mu I\end{aligned}\quad (5)$$

相关参数利用1stOpt拟合得到,利用Python中matplotlib库得出模型预测的疫情图如图6所示。

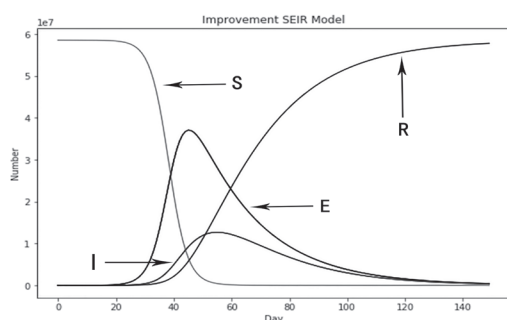


图6 疫情分布图

图5、图6,分别对应的是SEIR模型和改进的SEIR模型,在具有隔离仓室后,疫情爆发的时间延后了,爆发的规模和疫情高峰期也相应延后了,该传染病的潜伏者、易感者都显著减少,在只考虑隔离因素下,疫情的控制都取得了积极的作用,如果考虑政府实施的其他政策,如增加疫苗研发、媒体对新冠肺炎的宣传、提高传染病的治愈率等,就不难解释我国肺炎疫情从爆发到控制的短时高效。

4 结论

对SIR、SEIR及改进的SEIR模型进行比较,发现改进的SEIR模型考虑到了实际中传染病的传播,客观的表现了隔离对疫情传播的积极影响。中国能在极短的时间内控制疫情,保护人民生命财产安全,政府的干预调控起到了巨大的影响,

当今新冠肺炎在世界肆虐,其他国家应该学习我国对疫情的方针政策,尽快平息疫情,减少疫情带来的各种损失。

文中利用已有数据和软件,不能很好的表示疫情实际发生的时间段和各个人群所处疫情阶段的数量,主要原因有以下3点:(1)由于2月13日新冠肺炎确诊标准的改变,在拟合数据求取参数时将当日的病例数据去除了;(2)武汉交通管制时,虽然有潜伏者和感病者离开,但是绝大部分潜伏者和感病者还在武汉,而在数据预测时,用的是总的人口,没有考虑人口流出流入的变化,所以造成较大出入;(3)2月12日起,在政府的号召下全国医疗卫生及其他相关部门奔赴武汉支援,后期的隔离措施力度及治愈率应该对比前期更高,而在文中进行数据拟合时,没有将其纳入,所以拟合出的参数可信度不高,也对模型预测此次疫情趋势产生了不利影响。如果要精确预测全国疫情发展,对武汉“封城”前后的数据应该进行区分,对全国支援武汉前后也要进行区分,所以疫情的预测应该分成三个阶段,另外,此次新冠肺炎首先出现在武汉,湖北其他地区及其他各省份的预测应该基于辖区的实时数据,这些都是今后需要进一步研究的内容。

(责任编辑:侯辛锋)

参考文献:

- [1]Kermack W O,Mckendrick A G.Contributions to the Mathematical Theory of Epidemics[J].Proceedings of the Royal Society,1927,115 (05):700-721.
- [2]Ma Z, Jin Z. The stability of an SIR epidemic model with time delays[J]. Mathematical Biosciences and Engineering (Online), 2005, 3(1):101-109.
- [3] Guan W J,Ni Z Y,Hu Y,et al.Clinical characteristics of 2019 novel coronavirus infection in China[J].New England Journal of Medicine,2020.
- [4]邱明悦,胡涛,崔恒建.双区间删失下新冠病毒肺炎潜伏期分布的参数估计[J].应用数学期,2020,43(02):200-210.
- [5]余锦芬,宋玉凯,费菲,等.基于机器学习和动力学模型的湖北省新型冠状病毒肺炎疫情影响分析[J/OL].生物医学工程研究:1-15[2020-06-04].http://kns.cnki.net/kcms/detail/37.1413.R.20200422.1303.002.html.
- [6]朱连华,谭岩,肖惠文,等.基于阶段式传播模型COVID-19防控措施的量化评估分析[J/OL].南京信息工程大学学报(自然科学版):1-13[2020-06-04].http://kns.cnki.net/kcms/detail/32.1801.N.20200521.1428.002.html.

Prediction and Analysis of Novel Coronavirus Pneumonia Based on SIR Model

Kang Guanlong, Liu Bingxiang
(Jingdezhen Ceramic Institute,Jingdezhen 333403)

Abstract: This paper using python crawler technology, access to National Health and Family Planning Commission COVID released daily related cases (diagnosis, suspected, cured, death) data, research data, builds the epidemic trends, using the SEIR model and improved SEIR model to forecast the outbreak, and combined with the a series of interventions formulated by the government, to evaluate epidemic stages of development and trends, and affirmed the necessity of government intervention, so as to provide the references for the countries around the world to establish effective epidemic prevention and control.

Key words: Python crawler technology; SEIR model; New coronary pneumonia epidemic