新型冠状病毒传播的数学模型与预测

谢家荣^{1†}, 孙嘉辰^{2†}, 刘迅³, 孙小强⁴, 何群⁵, 胡延庆^{1*}

1. 中山大学数据科学与计算机学院, 广州 510006;

2. 中山大学电子与信息工程学院, 广州 510006;

3. 中山大学附属第三医院, 广州 510630;

4. 中山大学中山医学院, 广州 510080;

5. 广东省疾病预防控制中心/广东省公共卫生研究院, 广州 511430

† 同等贡献

* 联系人, E-mail: huyanq@mail.sysu.edu.cn

2020-02-18 收稿, 2020-04-07 修回, 2020-04-07 接受, 2020-04-08 网络版发表

国家自然科学基金(61773412, 61903385, 6177020844)资助

摘要 采用分阶段滚动SEIR疾病传播模型, 以国家卫生健康委员会公布的确证感染人数为基础, 测量了新型冠状病毒在人群中传播的基本再生数随时间的变化关系, 并根据该关系对新型冠状病毒肺炎(简称新冠肺炎)传染趋势进行了预测, 给出确诊人数每日的预测曲线. 预计疫情结束时全国除湖北省以外累计感染人数近14000人; 除武汉市之外, 累计感染人数近32000人. 另外, 全国非湖北地区, 在1月23日武汉封城后, 日新增长确诊人数在下降前是线性而非指数增加的, 这意味全国大部分地方对春节前夕从武汉流出的感染者的防控从一开始就是有效的.

关键词 新型冠状病毒, 分阶段滚动SEIR, 基本再生数, 传染病传播预测

2019年12月8日, 武汉市出现不明原因肺炎病例. 随后, 武汉市陆续出现多名不明原因肺炎患者. 2020年1月9日, 病原体被初步判断为新型冠状病毒; 1月22日, 湖北省启动突发公共卫生事件二级应急响应, 全国多地陆续出现新型冠状病毒肺炎患者. 截至2020年2月11日, 全国累计报告确诊病例44653例, 疑似16067例, 死亡1113例, 治愈4740例. 目前疫情发展形势仍十分严峻, 且确诊人数仍在不断上升, 严重影响了社会的稳定, 对人民群众生命安全和身体构成了重大威胁. 因此, 对疫情暴发与衰退进行准确预测以及对防范效率进行准确测量显得尤为重要, 它将为社会各级政府开展疫情防控, 以及人们日常工作与生活重新走上正轨提供重要的时间参考表, 具有重要的现实意义.

在当前信息快速传播的社会中, 在传染病传播周

期内, 人群一般都会经历“不重视-自我保护”2个阶段: 在第一阶段, 由于群众对疾病具体情况不知情, 对疫情的认知程度有限, 对疾病传播缺乏有效的防护措施. 在此阶段中的疾病传播过程可认为符合基本再生数为常数的经典Susceptible-Exposed-Infectious-Recovered (SEIR)模型^[1]. 在第二阶段, 由于政府和群众高度重视, 各种防治措施与资源被相继引入^[2], 加上对疾病传播途径已有较为准确的认知, 人群会逐渐采取有效且科学的防控措施, 使得基本再生数不断下降, 直到其下降到小于1, 并持续一段时间, 最终达到对疾病的根除.

本文在平均场近似下^[3], 即不考虑随机波动, 利用分阶段滚动SEIR模型, 基于国家卫生健康委员会公布的确证人数, 对新型冠状病毒肺炎(简称新冠肺炎)基本再生数进行了滚动测量. 经典SEIR建模方式为, 在给定再生

引用格式: 谢家荣, 孙嘉辰, 刘迅, 等. 新型冠状病毒传播的数学模型与预测. 科学通报, 2020, 65: 2348–2355

Xie J R, Sun J C, Liu X, et al. A predictive model for COVID-19 spreading (in Chinese). Chin Sci Bull, 2020, 65: 2348–2355, doi: [10.1360/TB-2020-0143](https://doi.org/10.1360/TB-2020-0143)

数为常数的基础上,处于I态的感染者由于被有效隔离而失去传播能力,他们相当于处于R态.因此,疾病传播主要依赖于处于已经被感染,但并未显示出病症的潜伏期(E态)的感染者.而本文使用的分阶段滚动SEIR模型与经典SEIR模型不同的是,滚动SEIR模型考虑了人群/政府采取有效防控措施下,基本再生数是随时间变小的.基于滚动SEIR模型,我们对再生数的下降趋势给出了预测,并基于再生数的预测曲线实现了对感染人数的预测.

1 传播过程的滚动SEIR数学模型

1.1 人群未采取任何措施情况下的SEIR传播模型

新冠肺炎基于人际接触的社交网络进行传播,在未采取措施的情况下,设每个感染者平均每天接触人数为 K .设感染者接触易感者后,以传播概率 β 感染易感者.则一个感染者平均每天会感染健康者的人数为 $k=\beta K$,基本自然再生数^[4-7] $R_0=kT$,其中 T 为潜伏期.在

简化情况下,我们采取平均场近似.这里忽略接触人数与社交网络的具体结构特性相关^[4,8-10].本文根据已知文献^[11],取 $T=5.2$.在下文对实际天数的讨论中,采用整数5天(方便计算,也讨论了6与7天的情况),每天感染能力一样不做区分.

基于上述模型的假定,设 $n(t)$ 为第 t 天的新增感染人数,设 $c(t)$ 为第 t 天的累计感染人数,则可列出迭代方程

$$\begin{cases} n(t) = [c(t-1) - c(t-T)]k, \\ c(t) = c(t-1) + n(t). \end{cases} \quad (1)$$

由此可得

$$c(t) = (k+1) \times c(t-1) - k \times c(t-T), \quad (2)$$

基本自然再生数 R_0 为

$$R_0 = \frac{n(t)}{c(t-1) - c(t-T)} T. \quad (3)$$

上述公式(1)是可解的:

$$c(t) = \begin{cases} m(k+1)^{t-1}, & 1 \leq t \leq T, \\ m[(k+1)^{t-1} - (t-T)k(k+1)^{t-T-1}], & T \leq t \leq 2T, \\ m\left[(k+1)^{t-1} - (t-T)k(k+1)^{t-T-1} + \frac{(t-2T)(t-2T+1)}{2}k^2(k+1)^{t-2T-1}\right], & 2T \leq t \leq 3T, \end{cases} \quad (4)$$

$$n(t) = \begin{cases} mk(k+1)^{t-2}, & 1 \leq t \leq T, \\ mk[(k+1)^{t-2} - (tk - Tk + 1)(k+1)^{t-T-2}], & T \leq t \leq 2T, \\ mk\left\{(k+1)^{t-2} - (tk - Tk + 1)(k+1)^{t-T-2} + \left[\frac{(t-2T)(t-2T+1)}{2}k + t - 2T\right]k(k+1)^{t-2T-2}\right\}, & 2T \leq t \leq 3T. \end{cases} \quad (5)$$

当 t/T 不是很大,并且 $(k+1)^T$ 较大时,公式(4)和(5)中起主要作用的是第一项:

$$\begin{cases} n(t) \approx mk(k+1)^{t-2}, \\ c(t) \approx m(k+1)^{t-1}. \end{cases} \quad (6)$$

2020年1月23日前后,据百度指数与Google搜索趋势显示,“武汉”、“新冠肺炎”、“新型冠状病毒”等关键词搜索热度急剧上升(图1(a), (b)).在同一天,武汉市政府开始限制市民离开武汉,并停摆市内公共交通,这意味着新冠肺炎在23日之后已经引起人们的广泛关注,群众开始意识到事态的严重性,开始进行有效的自我保护.当人群自我保护措施生效后,确诊感染人数速度就会变慢.因此,有理由相信,在1月23日的第5天,即1月28

日前,确诊感染人数大致是按自然增长方式增长的.此过程可用常数的基本再生数 R_0 ,即经典的SEIR模型进行拟合.对于实际感染人数的增长,我们收集了国家卫生健康委员会公布的从2020年1月15日到2月1日的全国确诊人数数据.由图1(c), (d), (e)可以看出,在1月28日前,确诊人数随天数的增长可以用公式(6)很好地拟合.

1.2 滚动SEIR传播模型

当人群采取各种有效的保护措施以规避疾病时,疫情传播就会开始衰减,如减少外出、聚会导致 K 降低;戴口罩、勤洗手则导致 β 降低.为了估计人群对新冠肺炎的反应情况,我们讨论人群采取有效自我保护措施后如何估计每天再生数 $\tilde{R}_0(t)$ 及单个感染者日均传

图1 (网络版彩色)新冠肺炎感染人数的指数增长。(a), (b) 为百度搜索指数与Google搜索指数。(c~e) 全国、全国除去湖北省、全国除去武汉市的累计确诊人数实际值与拟合值

Figure 1 (Color online) Exponential growth of the number of COVID-19 infections. (a) and (b) Baidu index and Google index of some related keywords, respectively; (c) predicted and ture accumulated number of confirmed infections in China. Similar curves in China expect Hubei Province and China expect Wuhan city are plotted in (d) and (e), respectively

染人数 $\tilde{k}(t)$ 随时间 t 的变化。 $\tilde{R}_0(t)$ 越小意味着大众自我保护的手段效果越好,当 $\tilde{R}_0(t)$ 小于1时,表示疾病得到了彻底的控制,持续一段时间后,疾病将会被根除。

与公式(1)相似,人群在自我保护的情况下,感染人数迭代方程为

$$\begin{cases} n(t) = [c(t-1) - c(t-T)]\tilde{k}(t), \\ c(t) = c(t-1) + n(t). \end{cases} \quad (7)$$

而新的再生数将是时间的函数:

$$\tilde{R}_0(t) = \frac{n(t)}{c(t-1) - c(t-T)} T. \quad (8)$$

基于国家卫生与健康委员会公布的截至2020年2月11日的确诊人数,我们对公式(8)进行了拟合,得到 $\tilde{R}_0(t)$ 如图2所示。从图2(a)可看出,基于全国数据计算的 $\tilde{R}_0(t)$ 在1月27日之前,基本上在3左右波动,其平均值可以近似地认为是自然基本再生数,即 $R_0=2.57$ 。

数据显示,武汉市日新增人数波动非常大(图2(f)),

特别是1月26~28日3天波动异常。同时,从大量相关报道来看,武汉市及其所在的湖北省医疗资源与人力可能比较紧缺,未必能对真实的已感染者做出快速检测与判断,导致日新增确诊人数可能偏小。所以在第一阶段(图2(c~e)),测量的自然基本再生数 $R_0=2.57$ (全国), $R_0=2.59$ (全国除湖北), $R_0=2.78$ (全国除武汉)的估计可能偏小。因此为了保险起见,我们分阶段计算了全国、全国除湖北省、全国除武汉市数据外再生数 $\tilde{R}_0(t)$ 的变化情况,如图2(a~c)所示,可以看出, $\tilde{R}_0(t)$ 在1月25日之前,波动比较大。而1月25日之后, $\tilde{R}_0(t)$ 呈现递减趋势,且在2月7日之后,疫情得到进一步稳定控制, $\tilde{R}_0(t)$ 在0.8~1.0之间波动。

另外,我们从图2(d~f)可以看到一个非常好的现象。春节前夕,也就是1月23日武汉封城后,日新增加确诊人数是线性的,而非之前的指数增长(图1(c~e))。这意味对春节前夕从武汉流出的感染者,全国大部分地区的防控从一开始就是有效的。

图2 (网络版彩色)滚动SEIR传播模型测算的 $\tilde{R}_0(t)$ 随时间的变化图。(a)~(c)分别为全国、全国除湖北省、全国除武汉市 $\tilde{R}_0(t)$ 变化。(d)~(f)分别为全国、全国除湖北省、武汉市新增确诊人数变化

Figure 2 (Color online) Value of $\tilde{R}_0(t)$ estimated by the stage-rolling SEIR model. (a-c) $\tilde{R}_0(t)$ of the whole country, the whole country except Hubei Province, the whole country except Wuhan city, respectively. (d-f) Plot the number of newly confirmed cases per day in the whole country, the whole country except Hubei Province and the whole country except Wuhan city, respectively

1.3 滚动SEIR模型对全国趋势的预测

下面我们基于滚动SEIR模型,对新冠肺炎疫情未来的感染日新增加确诊人数以及总感染人数进行预测。由于武汉市疫情严重,可能形成医疗资源与人力相对短缺的局面,导致公布的实际医院确诊的人数不一定能真实反映已经感染的人数。因此我们选择除武汉市外的数据进行预测,同时也对除湖北省以及湖北省除武汉市的预测进行讨论。我们按照公式(8)分别测算了全国除湖北省以外全国除武汉市以外的 $\tilde{R}_0(t)$ 的变化情况,分别如图3(a)和(b)所示。可以看出,从2月7日到11日,计算的 $\tilde{R}_0(t)$ 已进入平稳期,即基本在[0.8, 1]的范围内波动,我们有理由相信疫情已得到了控制。因为我们使用平稳期(2月7~11日,共5天)内 $\tilde{R}_0(t)$ 的平均值(表1)作为平稳值,来做最终感染人数的预测值。

另外,鉴于湖北省除武汉市之外的城市受疫情影响相对于武汉而言比较小,医疗资源相对比较充足,有理由相信,实际感染人数与诊断确诊人数比较吻合,这样也可以纳入到我们的模型预测范围内。我们利用滚动SEIR测算了湖北省除武汉市外的 $\tilde{R}_0(t)$ 。图3(c)显示,虽然 $\tilde{R}_0(t)$ 还没有出现明显的平稳趋势,但已明显显示

出进入了平稳期,同时与图3(b)的形态较为接近,故我们使用图3(b)得到的最后5天 $\tilde{R}_0(t)$ 平均值来预测湖北省除武汉市外感染人数。

综上,利用预测的平稳态的 $\tilde{R}_0(t)$ (表1)及公式(9),预测未来的增长变化情况:

$$\begin{cases} n(t) = [c(t-1) - c(t-T)] \frac{\tilde{R}_0(t)}{T}, \\ c(t) = c(t-1) + n(t). \end{cases} \quad (9)$$

将表1的 $\tilde{R}_0(t)$ 代入方程(10),预测了未来一个月的 $n(t)$ 和 $c(t)$,结果如图4(a~f)所示,在3月18日之后,全国除湖北省外的疫情大部分将结束,全国除湖北省以外累计感染人数近14000人。除武汉市外,累计感染人数近32000人。对湖北省内除武汉市,除了可用公式(9)进行预测外,也可用全国除湖北省外的预测值减去全国除武汉市外的预测值得到,即

$$\begin{cases} \bar{n}(t) = n_{wh}(t) - n_{hb}(t), \\ \bar{c}(t) = c_{wh}(t) - c_{hb}(t), \end{cases} \quad (10)$$

其中, $n_{wh}(t)$ 和 $n_{hb}(t)$ 分别表示全国除武汉市和全国除湖北省在第 t 天的日新增人数预测值($c_{wh}(t)$, $c_{hb}(t)$)同

图3 (网络版彩色)除去湖北省感染人数外的 $\tilde{R}_0(t)$ 实际测量与预测值. 实心圆点为实测值, 空心方点为预测值

Figure 3 (Color online) True and predicted values of $\tilde{R}_0(t)$ in the country except Hubei Province. The solid circle symbols denote the true value and the hollow diamond symbols denote the predicted one

表1 全国除去湖北与全国除去武汉地区的2月7~11日用不同的平均潜伏期测量的再生数的平均值

Table 1 Average basic reproduction number measured by various T in the country except Hubei Province and except Wuhan city between Feb. 7–11

区域	$T=5$	$T=6$	$T=7$
全国除去湖北	0.942	0.870	0.825
全国除去武汉	0.935	0.870	0.831

理). 分别用公式(9)和(10)对湖北省内除武汉市的确诊人数进行了预测, 如图4(e~f)所示, 可以看出, 两种预测结果相差不大, 湖北省内除武汉市最终累积感染人数接近18000人.

下面将从另外一个角度对我们的理论进行分析与验证. 在疾病传播的第一阶段, 假设是再生数为大于1的常数, 每天新增人数呈现出指数增长; 而在第二阶段, 假设再生数是小于1的常数, 每天新增人数呈现出指数衰减. 对此假设, 我们做了进一步验证. 分别对全国除湖北省、全国除武汉市、湖北省除武汉市的每天新确诊人数的数据, 在对数-线性坐标系下进行验证(图5). 可以看出, 在采取措施前(后), 新增确诊人数大致呈再生数为常数的指数增长(下降). 这两个阶段大概有一周的转化过程, 我们没有做讨论. 在第一阶段, 由公式(6)可得, 每天确诊人数大约是前一天的 $k+1$ 倍, 由此可知图5中斜率 α (见第一阶段)与 R_0 的应关系是

$$R_0 = kT = (10^\alpha - 1)T. \quad (11)$$

而后一阶段, 呈现出指数衰减时, 由公式(8)可得, 新的 α (见第二阶段)与 $\tilde{R}_0(t)$ 的应关系如下:

$$\begin{aligned} \tilde{R}_0(t) &= \frac{n(t)}{n(t-1) + n(t-2) + \dots + n(t-T+1)} T \\ &= \frac{1}{10^{-\alpha} + 10^{-2\alpha} + \dots + 10^{-(T-1)\alpha}} T. \end{aligned} \quad (12)$$

综上, 可以看出我们模型的主要假设和实际情况是吻合的.

2 讨论

根据我们模型的预测效果以及目前所观测到的日新增确诊病例数量, 已基本能够清楚观察到疾病的衰减速度. 政府管控的政策也持续保持稳定, 因此有理由相信, 这种衰减速度会持续到疾病结束. 我们也意识到, 本工作具有一定的局限性. 例如, 我们构建的滚动SEIR模型基于确诊报告日期. 从感染到出现症状, 再到确诊, 分别有一定的时间滞后. 同时, 从出现症状到报告确诊, 时间跨度受到前期检测能力与各地医疗资源的影响, 我们的模型未考虑此因素. 另外, 在估算1月28日之前, 即1月23日武汉封城再后推平均潜伏期5天, 湖北省以外地区的确诊病例以武汉市输出的病例为主, 湖北省其他城市也有武汉市输出病例. 我们未考虑武汉市输出病例的因素, 所计算的全国除湖北、全国除武汉基本再生数 $\tilde{R}_0(t)$, 肯定包含武汉市本身的影响. 这些都值得在未来数据充足的情况下进一步研究与讨论.

3 结论

本文采用滚动SEIR传播模型, 对新冠肺炎的传播进行建模, 并对其自然基本再生数和人群采取自我保护措施后再生数的变化情况进行了测量. 在此基础上

图 4 (网络版彩色)新冠传播预测人数. 圆点为实际值, 其余点为预测值. (a), (b) 全国除湖北省的预测数据. (c), (d) 全国除武汉市的预测数据. (e), (f) 湖北省除武汉市的预测数据

Figure 4 (Color online) Predicted number of COVID-19 infections. The circles are the true values while other symbols are the predicted values. (a) and (b) Predicted values in the whole country except Hubei Province; (c) and (d) predicted values in the whole country except Wuhan city; (e) and (f) predicted values in the Hubei Province except Wuhan city

图 5 (网络版彩色)新增确诊人数(对数坐标)的拟合斜率与基本再生数 $\tilde{R}_0(t)$, 分别为全国除湖北省(a)、全国除武汉市(b)、湖北省除武汉市(c)的数据. 测量数据之间的关系见公式(11), (12)

Figure 5 (Color online) Basic reproduction number $\tilde{R}_0(t)$ and fitted slope for the number of newly COVID-19 infections (in logarithmic coordinate). (a-c) Present the number of cases in the whole country except Hubei Province, the whole country except Wuhan city and the Hubei Province except Wuhan city, respectively. The relationship between $\tilde{R}_0(t)$ and fitted slope is described in Eqs. (11) and (12)

预测了新冠肺炎一个月内的日新增确诊人数,以及累计感染人数。预计全国除湖北省以外累计感染人数近14000人,除武汉市之外,累计感染人数近32000人。3月18日后新冠肺炎除湖北省以外,基本可以完全被根除。

另外,全国除湖北个别地方,在春节前夕,也就是1月23日武汉封城后,日新增加确诊人数是线性而非指数增加的,这意味对春节前夕从武汉流出的感染者,全国大部分地区的防控从一开始就是有效的。

参考文献

- 1 Lekone P E, Finkenstädt B F. Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 2006, 62: 1170–1177
- 2 Xie J R, Meng F H, Huang Y W, et al. Optimal devoted resource strategies to epidemic extinction by increasing recovery rate. *Intl J Mod Phys C*, 2019, 31: 1–10
- 3 Newman M E J. Spread of epidemic disease on networks. *Phys Rev E*, 2002, 66: 016128
- 4 Chowell G, Nishiura H. Transmission dynamics and control of Ebola virus disease (EVD): A review. *BMC Med*, 2014, 12: 196
- 5 Garnett G P. The basic reproductive rate of infection and the course of HIV epidemics. *AIDS Pat Care STDs*, 1998, 12: 435–449
- 6 Chen X L, Zhou T S, Feng L, et al. Nontrivial resource requirement in the early stage for containment of epidemics. *Phys Rev E*, 2019, 100: 032310
- 7 Dye C, Gay N. Modeling the SARS epidemic. *Science*, 2003, 300: 1884–1885
- 8 Watts D J, Strogatz S H. Collective dynamics of ‘small-world’ networks. *Nature*, 1998, 393: 440
- 9 Barabási A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286: 509–512
- 10 Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. *Phys Rev Lett*, 2001, 86: 3200
- 11 Li Q, Guan X H, Wu P, et al. Early transmission dynamics in wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med*, 2020, 382: 1199–1207

Summary for “新型冠状病毒传播的数学模型与预测”

A predictive model for COVID-19 spreading

Jiarong Xie^{1†}, Jiachen Sun^{2†}, Xun Liu³, Xiaoqiang Sun⁴, Qun He⁵ & Yanqing Hu^{1*}

¹ School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China;

² School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510006, China;

³ The Third Affiliated Hospital, Sun Yat-sen University, Guangzhou 510630, China;

⁴ Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China;

⁵ Guangdong Provincial Center for Disease Control and Prevention/Guangdong Provincial Institute of Public Health, Guangzhou 511430, China

† Equally contributed to this work

* Corresponding author, E-mail: huyanq@mail.sysu.edu.cn

As the novel coronavirus disease 2019 (COVID-19) has become a Public Health Emergency of International Concern, it is greatly significant to accurately predict the disease's incoming trend. Herein, we performed a stage-rolling Susceptible-Exposed-Infectious-Recovered (SEIR) model to measure the evolution of the basic reproduction number of COVID-19, based on the number of confirmed infections announced by the National Health Commission of the People's Republic of China.

We assumed that the infected number under the spreading of infectious diseases will generally experience two different stages. In the first stage, due to the public's ignorance of the severity and harmfulness of the disease, the infected population grows exponentially and the process of disease transmission can be considered as the classic SEIR model with a constant basic reproduction number. Consequently, with limited awareness of the epidemic situation, there is a lack of effective preventive measures to control disease transmission.

In the second stage, various control measures and medical resources are introduced in succession by the government, as well as the public gradually takes effective preventions (e.g. keep social distance and wear masks) based on the knowledge of the disease transmission. Collectively, the infected population grows much slower than the first stage. We performed a stage-rolling SEIR model, in which the basic reproduction number changes every day. Based on this model, the number of daily basic reproduction is estimated from the daily new infection number. We found that the daily basic reproduction number is expected to decline continually until it is less than 1, which means the eradication of the disease.

Leveraging the evolution of the basic reproduction number, we extrapolate the incoming daily basic reproduction number, based on which we further predict the incoming trend of COVID-19 spreading in terms of the daily infection number. Our predictive model estimates that at the end of the epidemic, the total number of infections in China is nearly 14000 except for Hubei Province, and 32000 except for Wuhan city. We also found that in most parts of China, the number of newly confirmed infections increases linearly rather than exponentially before the day of “Wuhan travel restrictions”, implying that the prevention and containing the infected people from Wuhan at the eve of the Spring Festival has been effective from the beginning.

COVID-19, stage-rolling SEIR, basic reproduction number, predicting epidemic tendency

doi: [10.1360/TB-2020-0143](https://doi.org/10.1360/TB-2020-0143)