

“归纳主义者的错觉”

基于元胞自动机的新冠病毒传播研究

高成志

张彬

李以煌

2020 年 7 月 10 日

摘要

新型冠状病毒疫情对全国造成了严重影响,人民的生产生活受到极大损失。应用数学建模方法,对病毒的传播过程和政府的防疫手段进行科学有效的定量分析与评估具有重要意义。

对于问题一,我们汇总了全国疫情新增、死亡、治愈人数的统计数据,应用 SEIR 模型对新冠疫情进行模拟,调整参数与实际数据进行拟合,并以此为基础得出了针对国内的病毒传播特征的统计描述。

对于问题二,我们将收集来的各省疫情发展情况的统计数据,运用批处理工具,生成各省疫情发展曲面图。通过对比,给出概括性的各省病毒传播特征结论。对各省疫情控制措施的有效性进行了定量分析。

对于问题三,我们结合美国疫情发展的统计数据,绘制了美国疫情发展趋势的时间序列图,通过分离各个关键参数,与中国疫情传播数据进行对比,找出两国疫情传播的异同,并试图分析其原因。

对于问题四,我们结合了美国六月之前的疫情发展情况,建立了元胞自动机模型进行模拟。再针对六月初的疫情数据,对模型的相关参数进行动态调优拟合,得出美国新冠感染人数在六月份的预测数据,最后与实际情况进行对比。

通过对比模型数据与实际数据我们发现,SEIR 模型对中国新冠病毒的传播趋势模拟较好。病毒扩散在短时间内得到控制充分反映出了中国防疫方案的先进性和有效性,为全球其他国家掌控干预节奏、制定防控策略提供了有益参考。而元胞自动机模型则难以预测变化多端的美国疫情发展,我们的预测结果过于乐观,与现实严重不符。针对这一现象我们进行了深刻的反思。

关键词: 新型冠状病毒; SEIR ; 元胞自动机; 动态调优; 数学建模

目录

1 问题背景与重述	5
1.1 问题背景	5
1.2 需解决的问题:	5
2 问题分析	5
3 模型假设	6
3.1 假设条件	6
3.2 符号说明	6
3.3 图表说明	7
3.4 数据来源 & 采集	7
3.5 国内病毒传播数据的统计分析	8
3.5.1 数据处理	8
3.5.2 SEIR 传染病模型的描述	10
3.5.3 模型实现	11
3.5.4 模型的调优与拟合	12
3.6 各省新冠传播情况的分析	14
3.6.1 数据处理	14
3.6.2 数据分析	14
3.7 美国新冠病毒传播数据的统计分析	16
3.7.1 数据处理	16
3.7.2 数据分析	17
3.8 美国新冠病毒传播的预测	18
3.8.1 元胞自动机模型的描述	18
3.8.2 模型实现	20
3.8.3 模型的调优与拟合	20
4 模型评价 & 改进	21
5 附录	25
5.1 声明	25
5.2 domestic.m	25
5.3 read.m	27

5.4	simulation_2.m	30
5.5	simulation.m	33
5.6	usa.m	34
5.7	map.m	36

1 问题背景与重述

1.1 问题背景

2020 年初，一场由新型冠状病毒引发的疫情在武汉迅速传播，并逐渐蔓延到全国，人民的生产生活受到了极其严重的负面影响。党中央领导高度重视，立即采取了强有力的防治措施，在各级人民政府的协同行动和广大人民群众积极配合之下，国内的疫情很快得到了初步的控制。

与此同时，新型冠状病毒继续向海外扩散。疫情在几个主要的西方国家（尤其是美国）并没有得到很好的控制。特别是进入六月以来，随着以美国为首的西方国家不断爆发大规模的反种族主义抗议游行，使得原本快要被遏制的疫情又有了反扑的可能。

截至北京时间 6 月 11 日 7 时 42 分，全球新型冠状病毒确诊病例达 7,347,323 例，死亡病例为 415,174 例。美国确诊病例达 1,997,636 例，死亡病例为 112,769 例。这些数字还在不断的增长中。

1.2 需解决的问题：

1. 分析国内的病毒传播数据（统计描述）。
2. 综合评估各省新冠传播的速度、控制措施的有效性。
3. 通过数据分析对比中国新冠传播与美国新冠传播的区别。
4. 预测美国新冠感染人数在六月份的数据。

2 问题分析

对于第一个问题，我们将汇总全国疫情新增、死亡、治愈人数的统计数据，使用 SEIR 模型对新冠疫情进行模拟，调整参数进行拟合，得出针对国内的病毒传播的统计描述。

对于第二个问题，我们将收集各省疫情发展情况的统计数据，运用批处理工具，生成各省疫情发展曲线进行对比，给出概括性的传播特征结论。对疫情控制措施的有效性进行定量分析。

对于第三个问题，我们将结合美国疫情发展的统计数据，对比中国新冠疫情的传播特征，找出两者的异同点，并尝试进行分析。

对于第四个问题，我们将结合六月之前的疫情增长情况，运用元胞自动机模型进行预测。再针对六月初的疫情数据，对模型的参数进行动态调优拟合，得出美国新冠感染人数在六月份的预测数据，最后与实际情况进行对比。

3 模型假设

3.1 假设条件

1. 假设收集的数据能客观公正地表现疫情的实际情况。
2. 感染系数 β 在封城后保持不变。
3. 患者的恢复系数 γ 保持不变。
4. 潜伏者的发病概率 α 保持不变。

3.2 符号说明

表 1: 符号说明

符号	解释
S	易感者 (Susceptible)
E	暴露者 (Exposed)
I	感病者 (Infectious)
R	康复者 (Recovered)
r	感染患者 (I) 每天接触的易感者数目
β	传染系数
α	潜伏者的发病概率
t	时间间隔的单位，默认为天

表 2: 图表说明

符号	解释
total	累计确诊人数
Increased	新增人数
InfectedNow	现有确诊人数
death	累计死亡人数
cure	累计治愈人数
HMT	港澳台累计确诊人数
abroad	境外累计确诊人数
positive	美国累计确诊
negative	美国检测后成阴性

3.3 图表说明

3.4 数据来源 & 采集

国内数据 clone 自 GitHub 相关数据收集仓库 [1], 使用 Perl 运用正则表达式进行数据剔除, 修正了不正确的 `xlsx` 数据格式。美国疫情数据来源于教师提供的 `dailyAmerican.xlsx` 表格和美国疾控中心 (CDC) 官网 [10] 的数据。

经过滤后国内数据的表头为:

时间 | 确诊人数总和 | 新增人数 | 现有确诊人数 | 死亡人数 | 治愈人数 | 港澳台共计 | 累计境外输入

各省数据表格的表头为:

Province | State | Confirmed | Death | Recovered

美国疫情数据的表头为:

date | states | positive | negative | pending | hospitalizedCurrently | hospitalizedCumulative | inIcuCurrently | inIcuCumulative | onVentilatorCurrently | onVentilatorCumulative | recovered | dateChecked | death | hospitalized | lastModified | total | totalTestResults | posNeg | deathIncrease | hospitalizedIncrease | negativeIncrease | positiveIncrease | totalTestResultsIncrease

3.5 国内病毒传播数据的统计分析

3.5.1 数据处理

首先我们依据处理后的表格 meta.xlsx 绘制出了中国新冠疫情发展的 85 天趋势图（数据统计开始于 2020 年 1 月 10 号，截止于 2020 年 4 月 3 号），代码见附录 domestic.m 文件。

我们可以很清楚的看到，在第 34 天，数据有了明显的异常（图 2）波动：

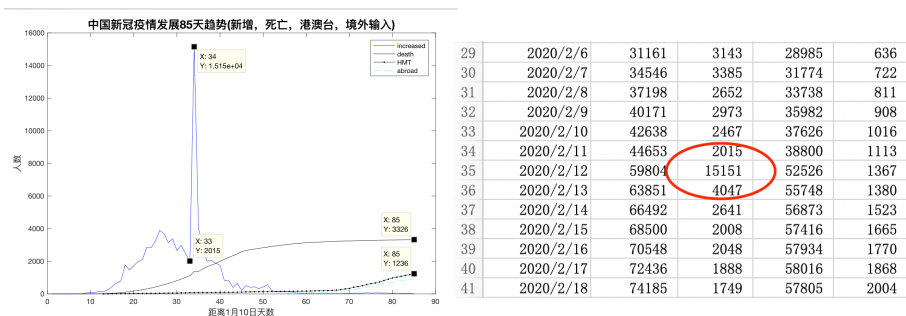


图 1: 中国新冠疫情发展趋势

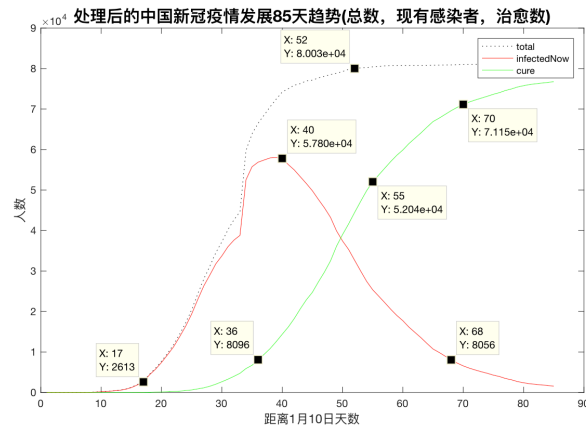
图 2: 数据变更说明图

这是由于当天疫情统计部门临时更换统计策略造成的。

为了更好地减少由于政策变更对统计数据造成的测量影响，我们将第 34、35 两天的数据做了平滑处理，处理后的数据依据数字数量级分为 A、B 两图重绘如下：

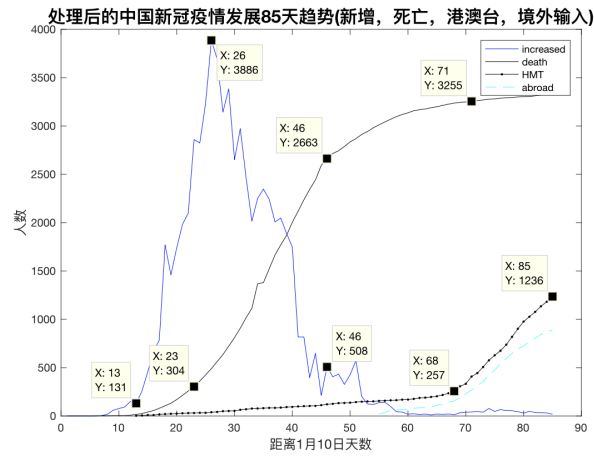
从 A 图可以看出，累计确诊人数在统计数据开始后的第 52 天（3 月 1 号）到达增长顶点（80,026 人），随后进入平台期，标志着国内疫情基本得到控制。

而现有确诊人数自统计第 17 天（1 月 26 日）以来迅速增长，到第 40 天（2 月 18 日）到达增长顶点（57,805 人），随后快速下降，到第 68 天开始下降速度开始放缓。康复人数的大幅增加的时间段正位于现有确诊人数快速下降的阶段。



[H]

图 3: 处理后的中国新冠疫情发展趋势 A



[H]

图 4: 处理后的中国新冠疫情发展趋势 B

从 B 图中我们可以看出, 新增人数从开始统计 (1 月 1 日) 时的大幅上升到从第 26 天 (2 月 4 日) 开始大幅下降, 整个转变过程非常剧烈, 并不符合传统病毒传播模型的一般曲线。

结合武汉的封城日期 (1 月 23 号) 我们推测: 这是由于疫情初期的病毒传播没有受到地方防疫部门的足够重视, 又恰好赶上春运的大规模人员流动, 给病毒急剧传播营造了有利条件, 因此有了非常陡峭的增长曲线。而之后的城市封锁又以一种极端的方式瞬间锁死了人员流量, 从源头切断了病毒传播, 造成了如图所示的急剧下降的新增人数曲线。

可以发现, 封锁交通在疫情初期的效果也并不是立竿见影的, 而是有一个过程, 大概在七天左右。

比较有意思的是到第 68 天 (3 月 17 日), 港澳台确诊和境外输入的病例明显开始上涨, 这标志着新冠病毒开始向外国扩散。关于这部分的内容分析我们将把它留在国外疫情部分。

接下来我们将对疫情传播进行数学建模。

3.5.2 SEIR 传染病模型的描述

SEIR 传染病模型是一种常见的病毒传播模型, 它将人群分为以下四个主要部分:

1. S 类人群, 指的是病毒易感人群 (Susceptible), 尚未感染病毒, 但缺乏病毒免疫能力, 与病毒感染者接触后容易感染病毒。
2. E 类人群, 指的是接触过感染者 I, 暂无能力传染其他人的类, 暴露 (Exposed) 在公共环境中, 适用于潜伏期较长的传染疾病。
3. I 类人群, 感染者 (Infections), 感染过病毒, 并有较大概率传播给 S 类人群, 将其转化成 I 或 E 类人群。
4. R 类人群, 康复或移除 (Removed or Recovered), 通常病人康复或死亡后均不具有明显感染能力, 故将该类人群移除出统计范围。

四者转化关系的参数为:

1. r 为感染患者 I 每天接触的易感人群的数目, 初始值设置为 30。
2. β 为传染系数, 暴露人群被感染的可能性, 初始值设置为 0.03。
3. γ 为恢复系数, 通常设置为康复时间的倒数, 这里设置成 0.1。

4. α 潜伏者的发病概率, 通常设置成患者潜伏期的倒数, 这里设置成 0.1。
5. t 为时间度量的离散间隔, 设置为 85 个单位 (天)。
6. N 为初始的全体人数。

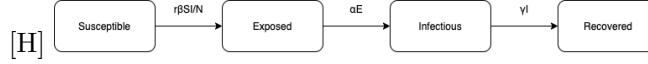


图 5: 转换流程图

转化的方式是:

$$\begin{aligned}
 \frac{dS}{dt} &= \frac{-r\beta IS}{N} \\
 \frac{dE}{dt} &= \frac{r\beta IS}{N} - \alpha E \\
 \frac{dI}{dt} &= \alpha E - \gamma I \\
 \frac{dR}{dt} &= \gamma I
 \end{aligned}$$

当 $dt=1$ 时, 上式的离散形式表现为:

$$\begin{aligned}
 S_{t+1} - S_t &= \frac{-\gamma\beta I_t S_t}{N} \\
 E_{t+1} - E_t &= \frac{\gamma\beta I_t S_t}{N - \alpha E_t} \\
 I_{t+1} - I_t &= \alpha E_t - \gamma I_t \\
 R_{t+1} - R_t &= \gamma I_t
 \end{aligned}$$

3.5.3 模型实现

据此我们编写了 matlab 程序 (代码见附录 simulation.m 部分) 来展示 SEIR 模型模拟下的病毒传播趋势 (图 6)。

这是一个非常漂亮的理论模拟曲线, 但是它有一些问题:

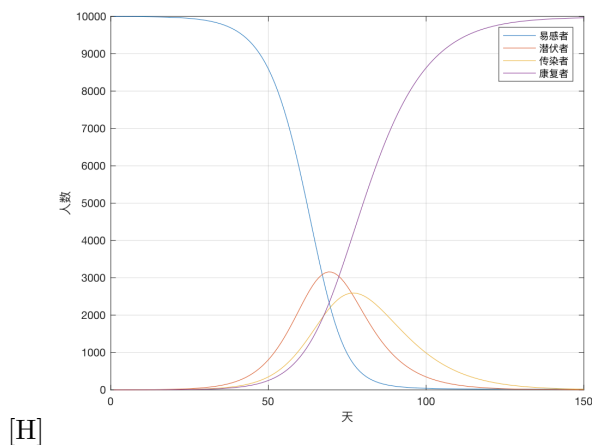


图 6: SEIR 模型的病毒传播趋势

1. 全国疫情的重灾区是武汉地区，而武汉是一座有着 1400 万人口 [2] 的大都市，本模型的初始人口数据明显偏小了，为后面的曲线拟合带来了麻烦。
2. 模型的增长曲线非常光滑，这是由于没有考虑封城给病毒传播带来的影响造成的，这不符合实际新冠病毒的发展趋势。
3. 由于人口基数过小，最终被感染的人数的比例明显偏高了，人口调整后之后总人口在数值上变化不明显，对后期比较是一个干扰项。

针对以上的种种问题，我们修改了程序（代码见附录 simulation2.m 部分）增加了初始人数，设置了封城的触发条件，调节了相关参数，重新绘制了模拟曲线图。（图 7）

这与我们之前看到的急剧的增长曲线非常相像，接下来我们将会针对特定参数进行调优，并与实际情况进行拟合。

3.5.4 模型的调优与拟合

关键参数的调整落在了感染患者 (I) 每天接触的易感人群的数目 r 、传染系数 β 和隔离条件上的设置上。其中 r 控制着波峰的高度， β 控制前期的增长坡度，隔离条件控制着下降的坡度。

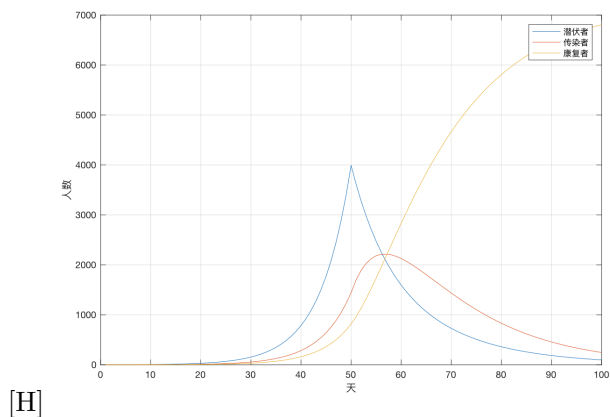


图 7: 修正后的 SEIR 模型的病毒传播趋势

经过一定的调整，我们发现最合适的 r 为 25， β 为 0.03，隔离条件设置为 r 降到 1（人）。

将图 4 的新增曲线离散化，还原数据集，并将其贴在图 7 中进行对比：

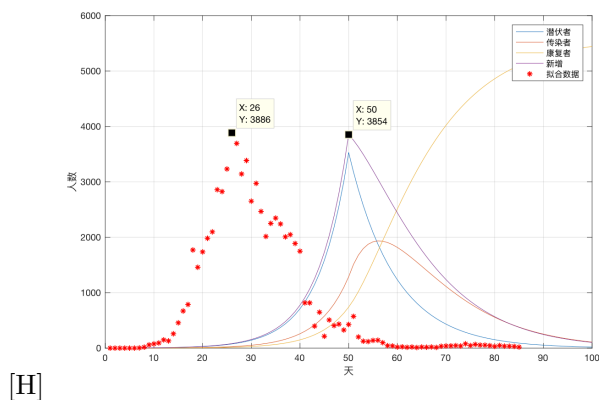


图 8: SEIR 模型的病毒传播趋势 (拟合图)

紫色的新增曲线是潜伏者和传染者的加权和（有未被检测可能）。可以看到新增人数和拟合数据匹配得非常好。同时我们可以看到两个波峰之间有一个时间差，大约是 24 天，这说明疫情从刚刚开始到开始统计（1 月 10 号）已经过去了 20 多天，恰好符合疫情 12 月中旬开始的实际情况。

为了更好的比对数据，我们将离散数据集右移 $24t$ 个单位，重新绘制曲线。

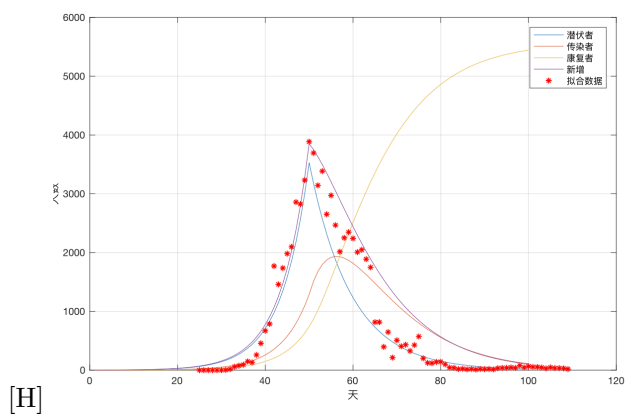


图 9: 移动后的 SEIR 模型的病毒传播趋势 (拟合图)

可以看出，模型对疫情发展实际趋势拟合良好。

3.6 各省新冠传播情况的分析

3.6.1 数据处理

问题主要集中在各省 Excel 表格的数据上。

由于数据全被集中到了一个表中，而 Matlab 的原生表格读取语句只能一次性读取一个表格的第一个工作簿，不得不使用 visual basic 语言编程将表格进行了拆分。

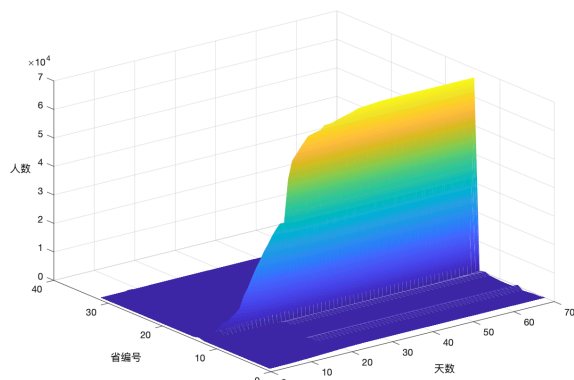
紧接着我们发现每省的排序的顺序有问题的：部分表格被按照感染人数排序，这样就没有办法反映出各省增长的连续情况。我们使用人工排序的方法，将 69 个表格按照各省首字母进行排序。

在使用 MATLAB 编程时发现，由于疫情前期部分省市数据有缺失造成矩阵维数不一致，主要是新加入统计的省份挤占了其他省份的位置。还有不少多报漏报的现象，我们全部进行了手工校正。

3.6.2 数据分析

依据 34 个省的 69 天连续统计数据，我们绘制出去除网格的表面图如下：

从图 10 可以看出，湖北省的数据太突出，抹平了其他地区的数据差异，



[H]

图 10: 全国各省疫情发展统计图 (确诊)

为了方便观察应该去除。去除之后的统计清晰地显示了各省之间的防疫差异。

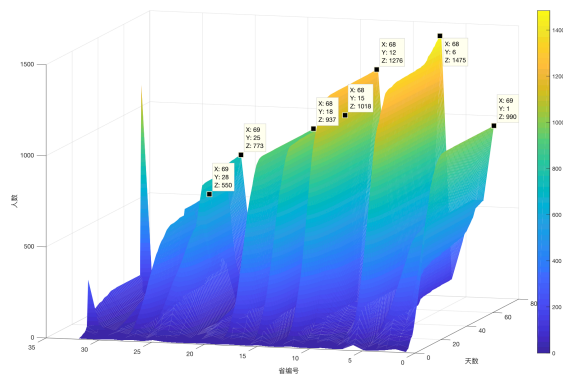
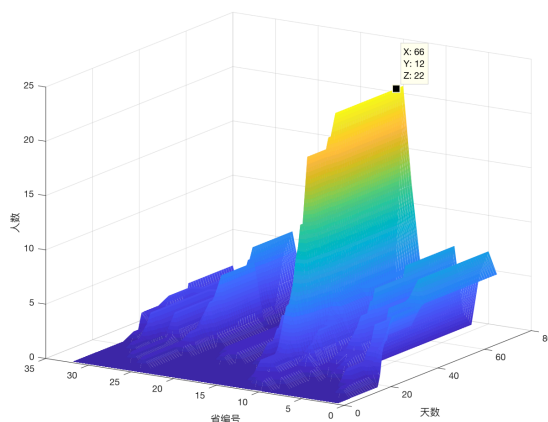


图 11: 全国各省疫情发展统计图 (累计确诊)

各省累计确诊的增长曲线都比较相近，并不能明显地区分当地的防疫效果的优劣。其中广东（6）与河南（12）在人数上最为突出，广东在统计的后段还出现了疫情反扑的趋势，可能是由于港澳台和境外输入人口的缘故。除此之外各省防疫措施从结果来看都没有表现出明显差异，感染人数多少主要与人口流量和人口密度决定。

从图 12 可以看出，在累计死亡人数上，河南省的数据要远远出其他省



[H]

图 12: 全国各省疫情发展统计图 (死亡)

份。

有人 [3] 分析认为这是由于初期感染基数较大造成的：

“河南省 2020 年 2 月 16 日感染确诊人数在非湖北省排名第二仅次于广东省，但是河南第一代感染基数大。地理位置上河南和湖北相邻，郑州又是全国交通枢纽。河南省作为一个外出务工人员大省，有很多从湖北返乡务工人员，由于湖北及全国官方通报延迟导致隔离确诊措施延迟，二代三代患者迅速生成相继发病。一代重症患者没有得到及时隔离救治，一二代轻症患者发展成重症患者造成救治困难。”

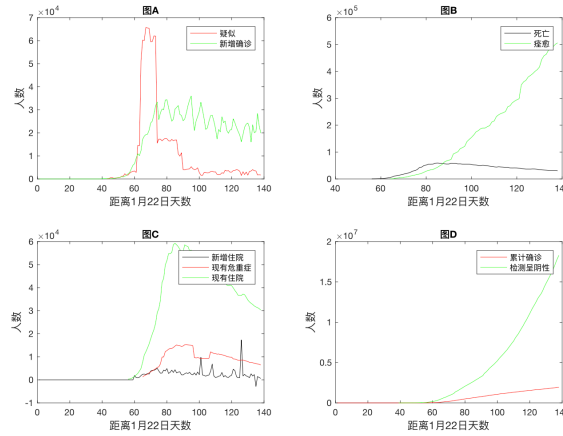
事实上，即便是死亡人数最多的河南省也仅仅有 22 人，随机事件对数据影响的参数更大，并不能得出河南省防疫不力的结论。

3.7 美国新冠病毒传播数据的统计分析

3.7.1 数据处理

我们首先将 dailyAmerican.xlsx 中的数据进行处理，删去不必要的项，然后编写 matlab 程序（代码见附录 usa.m 部分）绘制时间序列图如下：

（由于各统计项的数量差距较大，我们将统计数据分成 A、B、C、D 共 4 个子图呈现。）



[H]

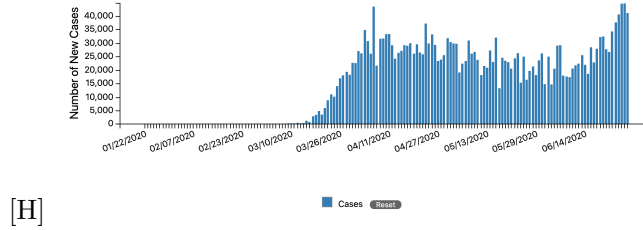
图 13: 美国疫情发展统计图

3.7.2 数据分析

从图 A 可以看出, 新增确诊人数几乎是从第 50 天 (3 月 11 日) 一夜之间上涨上来的, 有人认为这是美国政府在疫情前期检测量不足造成的统计误差, 但我们可以通过比对国内统计数据发现, 美国新冠病毒爆发式增长的时间恰好和图 5 中国统计境外输入病例的快速上涨时间重合。鉴于此时的中国已经建立了较为完善的病毒检测体系, 因此这种说法是不准确的。

从全部 A、B、C、D 四张图中都可以发现病毒传播速度在第 60 天 (3 月 21 日) 的增长达到了高峰, 到第 80 天增长到达极限, 该过程一共 20 天左右, 这恰好和图 4 中中国疫情发展时间相同。之后美国新冠病毒新增数据趋于稳定, 而住院、死亡人数开始逐渐下降。

但值得注意的是, 与图 4、5 中国疫情发展趋势相比, 图 A 显示的每日新增人数在到达高峰之后并没有快速下降, 而是在高位不断波动。根据我们在美疾控中心 [10] 得到的后续疫情发展数据, 直到六月末, 疫情不仅没有能够快速得到遏制, 反而在有了抬头的趋势 (见图 14)



[H]

图 14: 美疾控中心每日新增数据

对于这一现象，哥伦比亚大学梅尔曼公共卫生学院环境卫生科学系发表的论文 *Projection of COVID-19 Cases and Deaths in the US as Individual States Re-open May 4, 2020* [11] 是这么解释的：

”In March and April 2020, control measures enforcing social distancing and restricting individual movement and contact were adopted across the United States in an effort to slow the spread and growth of COVID-19. However, a number of states have now begun to ease these restrictions.”[11] 可见，在疫情后半段显示出的两国病毒传播的发展趋势上的差异是由两国不同的防疫措施引起的：美国部分州在疫情还没能得到控制的情况下贸然解封，导致新增感染人数重新上升。

我们通过美疾控中心官网和纽约时报 [9] 相关报道可以找到美复工和停摆的州的动态图（图 15），通过比对可以发现，疫情高发州在短暂复工后又纷纷重新开始施行隔离措施。

3.8 美国新冠病毒传播的预测

3.8.1 元胞自动机模型的描述

由于美国疫情主要发生在纽约等大城市中，人员密集度高，而封锁又不彻底，病毒传播因素变化大。传统的病毒感染模型并不能很好的模拟这一动态情况。所以我们选择了元胞自动机模型来模拟美国疫情的传播情况。

细胞自动机（英语：Cellular automaton），又称格状自动机、元胞自动

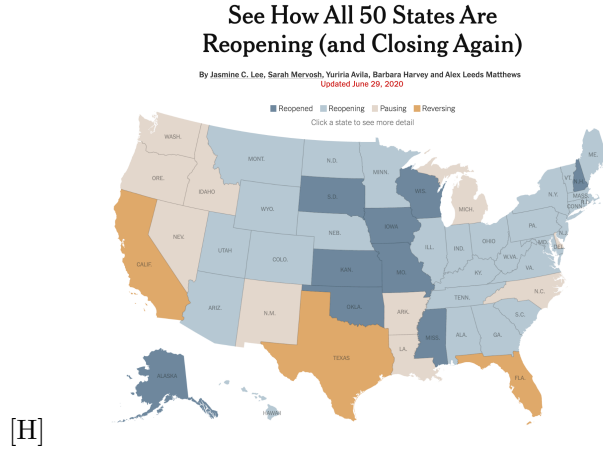


图 15: 纽约时报: 复工后又再度关闭的州

机，是一种离散模型，在可计算性理论、数学及理论生物学都有相关研究。

它是由无限个有规律、坚硬的方格组成，每格均处于一种有限状态。整个格网可以是任何有限维的。同时也是离散的。每格于 t 时的态由 $t-1$ 时的一集有限格（这集叫那格的邻域）的态决定。每一格的“邻居”都是已被固定的。（一格可以是自己的邻居。）每次演进时，每格均遵从同一规矩一齐演进。[4]

标准的元胞自动机 (A) 由元胞、元胞状态、邻域和状态更新规则构成。数学表示为 [5]。

$$A = (L, D, S, N_1, f)$$

其中 L 为元胞空间； d 为元胞自动机内元胞空间的维数； S 是元胞有限的、离散的状态集合； N_1 为某个邻域内所有元胞的集合； f 为局部映射或局部规则。元胞空间是元胞所分布的空间网点的集合。理论上元胞空间在各个维向上是无限延伸的，为了能够在计算机上实现，而定义了边界条件，包括周期型、反射型和定值型 [6]。一个元胞通常在一个时刻只能取自一个有限集合的一种状态，例如 0,1。元胞状态可以代表个体的态度，特征，行为等 [7]。在空间上与元胞相邻的细胞称为邻元，所有邻元组成邻域。

元胞自动机的一大魅力在于，尽管它本身的理论描述有着近乎唬人的复杂度，但它的实现却是非常简单直观的，它给了建模者相当大的自由度来

决定模型的实现，从最简单的用来演示的生命游戏，到妄图解释整个宇宙运行规则的万有理论 [12]

3.8.2 模型实现

我们依据元胞自动机原理编写了 matlab 程序（代码见附录 map.m 部分）我们将初始元胞大小设置为美国人口（截至 2019 年 1 月为 3.3 亿 [8]）的 $1/1000$ ，单位长为 $575 = \sqrt{330000}$ 。

为了简化模型，我们设置了两个元胞动态参数：人群感染概率（getInfected）和人员流动密度（peopleInGrowth），初始值分别为 0.06 和 0.005。分别由一个 575×575 的伪随机矩阵分发。

元胞自动机的维数为 3，第零层为 R 人群（removed）用黑色表示，第一层为 I 人群（infected）用红色表示，第二层为 S 人群（susceptible）用绿色表示。为了使模型模拟过程直观，我们采用了 MVC 设计模式将业务逻辑、数据和界面显示分离，将元胞矩阵的动态更新分别映射到一个平面像素图和一个折线图上，并实时刷新数据（图 16、17）。

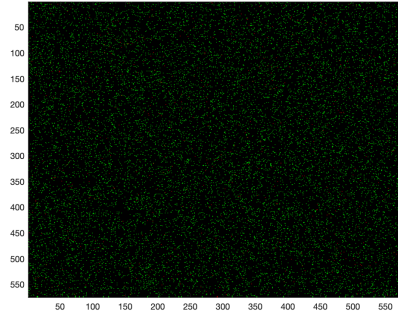


图 16: 元胞自动机模拟图

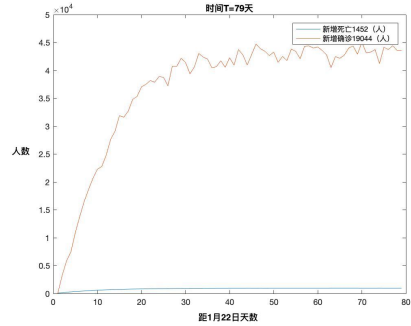


图 17: 元胞自动机折线动态图

3.8.3 模型的调优与拟合

静态的元胞参数显然没法反映变化多端的美国新冠疫情数据变化情况。为此，我们基于现实的美国新冠疫情数据对人群感染概率（getInfected）和

人员流动密度 (peopleInGrowth) 两个参数进行了动态调优。重绘后的折线图 and 实际数据拟合如下 (图 18)

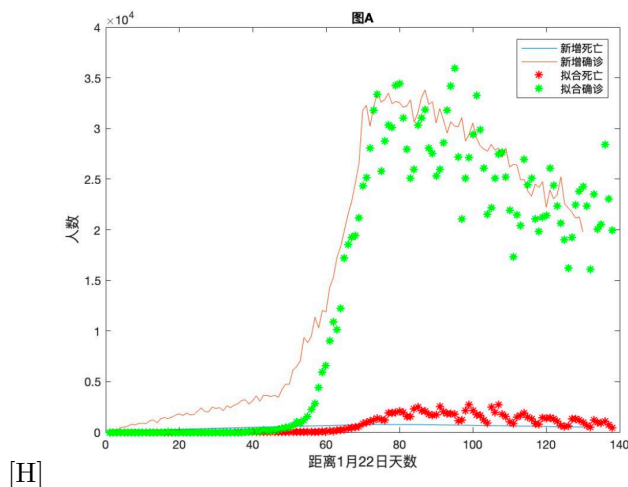


图 18: 元胞自动机折线拟合图

可以看出模型在后半段对疫情发展实际趋势拟合良好, 因此我们将元胞更新的时间周期延长至于 160 天 (6 月 30 日) 重新模拟来预测美国六月的疫情发展情况。

模拟结果显示, 在我们的模型中新增确诊人数会在六月末逐渐下降, 死亡人数则保持稳定。数据经过累加我们预测美国六月份感染总人数会达到 2,345,299 人。(图 19)

然而在本文写作的同时, 六月份逐渐过去, 现实情况却与我们预料的有所不同。

4 模型评价 & 改进

现实情况与我们预测的大相径庭, 美国疫情进入六月末之后感染人数快速上升, 不久就突破了 250 万, 多个州在重启之后再度封锁……

这不禁让人想起伯兰特·罗素著名的火鸡问题: 火鸡中的一个科学家经过一年的不断观察, 发现农场主每天早上 11 点的时候会给它们投饲料, 于是在感恩节这一天它宣布了自己这个伟大发现。但是早上 11 点过去, 农场

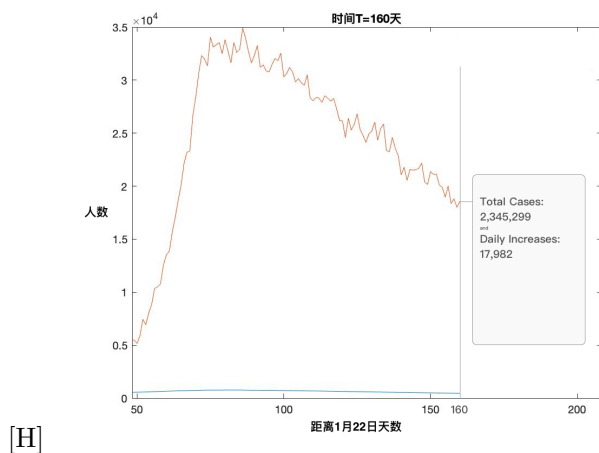


图 19: 美国疫情数据预测

主并没有来投饲料，反而带了一把枪把它们全杀了。

火鸡问题多多少少反映出了归纳主义者的困境。“过去的未来和未来的未来是否相似”这个问题不能由过去的经验解决，只有等未来已来时才能得到验证，但很多人误以为验证得来的结论就是真理，有的时候这不过是一种侥幸罢了。

在这个数学模型中。我们原以为增加伪随机函数分发的元素，可以在虚假的随机中找出不由随机事件变化的规律。但事实上世界的发展是可能是由无限不可列多个参数决定的（当然，这是一种陈旧的机械决定论的想法），谁也不知道未来会发生什么。基于过去的假设，通常在未来的突变来临时不堪一击。

现在人工智能和大数据的概念被媒体和金融投机者炒得火热，很多不知情的外行人把它吹捧成无所不能的万金油。事实上，所谓的人工智能和”智能“这个概念还相差很远，倒是和中世纪的炼金术很相像：将参数的设置和拟合过程由大量计算黑箱自动求解，通过已有的数据训练出针对未来情形的通用解决公式。这在客观条件变化有限的应用场景已经取得了突破性的进展，例如围棋和识别文字，但在更加复杂的情形却注定是无所作为的，原因就在于计算机不能理解无限，它所能做的只不过是有限的位运算来笨拙的模拟无限的概念。

回到本模型上，由于编程人员的懒惰，大多数结论性的评价没有通过统

计学的计算语言表达，而仅仅通过直观反映。这是本文最需要改进的地方，但是由于三位作者的统计学水平近乎文盲，直到写作完成这个问题也未能得到解决，成了一个不小的遗憾。

在附录里，您很容易发现源代码混乱不堪，Debug 处理和多功能的注释混杂在一起。尽管我们打算在编写附录时重新整理代码，但是您依然可能发现在您的计算机中无法正常运行文中展示的全部功能。

本文的写作者在中途因为不堪忍受 word 在处理中型文档时的频繁卡顿和崩溃，将写作环境迁移到了 L^AT_EX 上。因此论文格式的规范可能是本文为数不多的一个优点，但这完全不是我们的功劳。

在参数的拟合上，特别是元胞自动机参数的动态拟合上。原本打算采用 CoreML 中较为成熟的机器学习框架实现，但是由于本文的案例较为简单，而 CoreML 的 API 文档时着实较为复杂。最后改成了单位花费最小的人工实现，希望能在以后的学习中掌握更多对数据进行操作的工具和方法，写出一篇真正像样的论文。

(完)

参考文献

- [1] <https://github.com/BlankerL/DXY-COVID-19-Data>
- [2] https://web.phb123.com/city/renkou/city_241.html?ivk_sa=1023345q
- [3] <https://www.zhihu.com/question/372225552/answer/1019781435>
- [4] <https://w.upupming.site/wiki/元胞自动机>
- [5] S. Amoroso; Y.N. Patt. Decision procedures for surjectivity and injectivity of parallel maps for tessellation structures. Journal of Computer and System Sciences. October 1972
- [6] 周成虎; 孙战利谢一春. 地理元胞自动机研究. 北京: 科学出版社. 2000: 26-51. ISBN 9787030081209.
- [7] Wolfram 2002, 第 231 页
- [8] <https://w.upupming.site/wiki/美国>
- [9] <https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html>
- [10] <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html>
- [11] Teresa Yamana; Sen Pei; Sasikiran Kandula; Jeffrey. Department of Environmental Health Sciences, Mailman School of Public Health, Columbia University. ShamanProjection of COVID-19 Cases and Deaths in the US as Individual States Re-open May 4, 2020
- [12] <https://www.wolframphysics.org/technical-introduction/potential-relation-to-physics/>

5 附录

5.1 声明

本文的 git 本地仓库由于体积原因已于写作完成之时全部从本机移除，但是您还可以通过访问该仓库在 GitHub 上的备份：

https://github.com/Gaochengzhi/math_module_task

来获取全部的文档、源代码、图片和表格。

5.2 domestic.m

```
1      clc; clear; close all;
2      % 中国感染数据处理
3      [data, text] = xlsread('meta.xlsx');
4      total = data(:, 2);
5      increased = data(:, 3);
6      infectedNow = data(:, 4);
7      death = data(:, 5);
8      cure = data(:, 6);
9      HMT = data(:, 10);
10     abroad = data(:, 11);
11     x = [1:85];
12     % 修正数据前
13     figure(1);
14     plot(x, total, 'k:', x, infectedNow, 'r', x, cure,
15          'g');
16     xlabel('距离1月10日天数', 'FontSize', 12);
17     ylabel('人数', 'FontSize', 12);
18     title('中国新冠疫情发展85天趋势(总数, 现有感染者,
19           治愈数)', 'FontSize', 16);
20     legend('total', 'infectedNow', 'cure');
21     figure(2);
22     plot(x, increased, 'b', x, death, 'k', x, HMT, 'k
23          .-', x, abroad, 'c--');
```

```
21 xlabel('距离1月10日天数', 'FontSize', 12);
22 ylabel('人数', 'FontSize', 12);
23 title('中国新冠疫情发展85天趋势(新增, 死亡, 港澳
      台, 境外输入)', 'FontSize', 16)
24 legend('increased', 'death', 'HMT', 'abroad');
25 % 修正数据后
26 [data, text] = xlsread('remeta.xlsx');
27 total = data(:, 2);
28 increased = data(:, 3);
29 infectedNow = data(:, 4);
30 death = data(:, 5);
31 cure = data(:, 6);
32 HMT = data(:, 10);
33 abroad = data(:, 11);
34 x = [1:85];
35 figure(3);
36 plot(x, total, 'k:', x, infectedNow, 'r', x, cure,
      'g');
37 xlabel('距离1月10日天数', 'FontSize', 12);
38 ylabel('人数', 'FontSize', 12);
39 title('处理后的中国新冠疫情发展85天趋势(总数, 现有
      感染者, 治愈数)', 'FontSize', 16)
40 legend('total', 'infectedNow', 'cure');
41 figure(4)
42 plot(x, increased, 'b', x, death, 'k', x, HMT, 'k
      .-', x, abroad, 'c--');
43 xlabel('距离1月10日天数', 'FontSize', 12);
44 ylabel('人数', 'FontSize', 12);
45 title('处理后的中国新冠疫情发展85天趋势(新增, 死
      亡, 港澳台, 境外输入)', 'FontSize', 16)
46 legend('increased', 'death', 'HMT', 'abroad');
```

5.3 read.m

```
1      clc; clear; close all;
2      files = dir('Provice/*.xlsx');
3      countNumber = [];
4      x = [1:69];
5      y = [1:32];
6      [xx, yy] = meshgrid(x, y);
7
8      for index = 1:69
9          filename = files(index).name;
10         index = num2str(index);
11         name = strcat('Provice/', index, '.xlsx');
12
13         % name = strcat('Provice/', filename);
14         [data, text] = xlsread(name);
15         countNumber = [countNumber, data(1:32, 1)];
16         z = countNumber;
17     end
18     figure(1)
19     surf(xx, yy, z)
20     shading interp;
21     xlabel('天数');
22     ylabel('省编号');
23     zlabel('人数');
24     %去除武汉
25     countNumber = [];
26     x = [1:69];
27     y = [1:32];
28     [xx, yy] = meshgrid(x, y);
29     for index = 1:69
30         filename = files(index).name;
31         index = num2str(index);
32         name = strcat('Provice/', index, '.xlsx');
```

```
33         [data, text] = xlsread(name);
34         countNumber = [countNumber, data(1:32, 1)];
35         z = countNumber;
36     end
37     z(14, :) = 0;
38     figure(2)
39     surf(xx, yy, z)
40     shading interp;
41     xlabel('天数');
42     ylabel('省编号');
43     zlabel('人数');
44     %死亡
45     countNumber = [];
46     x = [1:69];
47     y = [1:32];
48     [xx, yy] = meshgrid(x, y);
49     for index = 1:69
50         filename = files(index).name;
51         index = num2str(index);
52         name = strcat('Provice/', index, '.xlsx');
53
54         % name = strcat('Provice/', filename);
55         [data, text] = xlsread(name);
56         countNumber = [countNumber, data(1:32, 2)];
57         z = countNumber;
58     end
59     z(14, :) = 0;
60     figure(3)
61     surf(xx, yy, z)
62     shading interp;
63     xlabel('天数');
64     ylabel('省编号');
65     zlabel('人数');
```

```
66      %康复
67      countNumber = [];
68      x = [1:69];
69      y = [1:32];
70      [xx, yy] = meshgrid(x, y);
71      for index = 1:69
72          filename = files(index).name;
73          index = num2str(index);
74          name = strcat('Provice/', index, '.xlsx');
75
76          %      name = strcat('Provice/', filename);
77          [data, text] = xlsread(name);
78          countNumber = [countNumber, data(1:32, 3)];
79          z = countNumber;
80      end
81      z(14, :) = 0;
82      figure(4)
83      surf(xx, yy, z)
84      shading interp;
85      xlabel('天数');
86      ylabel('省编号');
87      zlabel('人数');
```

5.4 simulation_2.m

```
1      clc; clear; close all;
2      %人数
3      N = 15000000;
4      %暴露者
5      E = 0;
6      %感染者初值
7      I = 1;
8      %病毒易感人群
9      S = N - I;
10     %康复或移除人群
11     R = 0;
12     %感染患者I 每天接触的易感人群的数目
13     r = 25;
14     %传染系数
15     B = 0.03;
16     %潜伏者的发病概率
17     a = 0.1;
18     %恢复系数
19     y = 0.1;
20     %日期
21     T = 1:100;
22     [data, text] = xlsread('remeta.xlsx');
23     total = data(:, 2);
24     increased = data(:, 3);
25     infectedNow = data(:, 4);
26     death = data(:, 5);
27     cure = data(:, 6);
28     HMT = data(:, 10);
29     abroad = data(:, 11);
30     %平移前
31     %x = [1:85];
32     %平移后
```

```

33     x=[25:109];
34     for i = 1:length(T) - 1
35         %隔离条件
36         if i == 50
37             r = 0.1;
38         end
39
40         S(i + 1) = S(i) - r * B * S(i) * I(i) / N(1);
41         E(i + 1) = E(i) + r * B * S(i) * I(i) / N(1) -
            a * E(i);
42         I(i + 1) = I(i) + a * E(i) - y * I(i);
43         R(i + 1) = R(i) + y * I(i);
44         IN = E+I;
45         df1(i) = diff(IN,i);
46     end
47
48     plot(T, E, T, I, T, R,T,0.8*(I+E)); grid on;
49     xlabel('天'); ylabel('人数')
50     %拟合
51     %P=polyfit(x,cure',8);
52     %xi=0:.2:10;
53     %yi=polyval(P,xi);
54     %plot(xi,yi,x,cure','r*');
55     hold on;
56     %拟合
57     plot(x, increased', 'r*');
58     hold on;
59     legend('潜伏者', '传染者', '康复者','新增','拟合数
        据')
60     % plot(x, total, 'k:', x, infectedNow, 'r', x,
        cure, 'g');
61     % xlabel('距离1月10日天数', 'FontSize', 12);
62     % ylabel('人数', 'FontSize', 12);

```

```
63 % title('处理后的中国新冠疫情发展85天趋势(总数, 现
    有感染者, 治愈数)', 'FontSize', 16)
64 % legend('total', 'infectedNow', 'cure');
65
66 % figure(4)
67 % plot(x, increased, 'b', x, death, 'k', x, HMT, '
    k.-', x, abroad, 'c--');
68 % xlabel('距离1月10日天数', 'FontSize', 12);
69 % ylabel('人数', 'FontSize', 12);
70 % title('处理后的中国新冠疫情发展85天趋势(新增, 死
    亡, 港澳台, 境外输入)', 'FontSize', 16)
71 % legend('increased', 'death', 'HMT', 'abroad');
```


5.5 simulation.m

```

1      clc; clear; close all;
2      %人数
3      N = 100000000;
4      %暴露者
5      E = 0;
6      %感染者初值
7      I = 1;
8      %病毒易感人群
9      S = N - I;
10     %康复或移除人群
11     R = 0;
12     %感染患者I 每天接触的易感人群的数目
13     r = 20;
14     %传染系数
15     B = 0.03;
16     %潜伏者的发病概率
17     a = 0.1;
18     %恢复系数
19     y = 0.1;
20     %日期
21     T = 1:150;
22     for i = 1:length(T) - 1
23         S(i + 1) = S(i) - r * B * S(i) * I(i) / N(1);
24         E(i + 1) = E(i) + r * B * S(i) * I(i) / N(1) -
                a * E(i);
25         I(i + 1) = I(i) + a * E(i) - y * I(i);
26         R(i + 1) = R(i) + y * I(i);
27     end
28     plot(T, S, T, E, T, I, T, R); grid on;
29     xlabel('天'); ylabel('人数')
30     legend('易感者', '潜伏者', '传染者', '康复者')

```

5.6 usa.m

```
1      clc; clear; close all;
2      % 美国感染数据表
3      [data, text] = xlsread('dailyAmerican.xlsx');
4      %累计确诊
5      positive = data(:, 3);
6      %累计不确诊
7      negative = data(:, 4);
8      %疑似
9      pending = data(:, 5); %60000
10     %现有住院
11     hospitalizedCurrently = data(:, 6); %40000
12     %现有危重症患者
13     inIcuCurrently = data(:, 8); %10000
14     %痊愈
15     recovered = data(:, 12);
16     %死亡
17     death = data(:, 14);
18     deathIncrease = data(:, 20);
19     hospitalizedIncrease = data(:, 21);
20     negativeIncrease = data(:, 22);
21     positiveIncrease = data(:, 23);
22     death = data(:, 6);
23     x = [1:138];
24     subplot(2, 2, 1)
25     plot(x, pending, 'r', x, positiveIncrease, 'g');
26     xlabel('距离1月22日天数', 'FontSize', 12);
27     ylabel('人数', 'FontSize', 12)
28     legend('疑似', '新增确诊');
29     title('图A');
30     subplot(2, 2, 2)
31     plot(x, death, 'k', x, recovered, 'g');
32     xlabel('距离1月22日天数', 'FontSize', 12);
```

```
33     ylabel('人数', 'FontSize', 12)
34     legend('死亡', '痊愈');
35     title('图B');
36     subplot(2, 2, 3)
37     plot(x, hospitalizedIncrease, 'k', x,
           inIcuCurrently, 'r', x, hospitalizedCurrently, 'g
           ');
38     xlabel('距离1月22日天数', 'FontSize', 12);
39     ylabel('人数', 'FontSize', 12)
40     legend('新增住院', '现有危重症', '现有住院');
41     title('图C');
42     subplot(2, 2, 4)
43     plot(x, positive, 'r', x, negative, 'g');
44     xlabel('距离1月22日天数', 'FontSize', 12);
45     ylabel('人数', 'FontSize', 12)
46     legend('累计确诊', '检测呈阴性');
47     title('图D');
```

5.7 map.m

```
1      close;
2      clear;
3      clc;
4      n = 575; %元胞矩阵大小
5      getInfected = 0.06;
6      peopleInGrowth = 0.005;
7      LongDistanceRun = 0.005;
8      ShortDistanceRun = 0.003;
9      UpAndLeft = [n, 1:n - 1];
10     DownAndRight = [2:n, 1];
11     tableOfArea = zeros(n, n); %初始化
12     TimeInfectedCell = zeros(n, n);
13     % 数值维度规则
14     % removed == 0 黑
15     % infected == 1 红
16     % susceptible == 2 绿
17     imh = image(cat(3, tableOfArea, tableOfArea,
18                     tableOfArea));
19     m = ...
20         annotation('textbox', ...
21                     [0.1, 0.1, 0.1, 0.1], ...
22                     'LineStyle', ...
23                     '-', ...
24                     'LineWidth', ...
25                     1, ...
26                     'String', ...
27                     '123');
28     timeSetter = 0;
29     %反向学习参数
30     for i = 1:160
31         if i <= 40
```

```

32         getInfected = 0.01;
33         peopleInGrowth = 0.001;
34
35     end
36
37     if i >= 40 & i <= 70
38         getInfected = 0.00005 * ((i - 40)^2) +
39             0.01;
40         peopleInGrowth = 0.000003 * ((i - 40)^2) +
41             0.001;
42
43     end
44
45     if i >= 70
46         getInfected = 0.06;
47         peopleInGrowth = 0.002 + 0.002 * (150 - i)
48             / 80;
49     end
50
51     sum = (tableOfArea(UpAndLeft, :) == 1) ...
52         + (tableOfArea(:, UpAndLeft) == 1) ...
53         + (tableOfArea(DownAndRight, :) == 1) ...
54         + (tableOfArea(:, DownAndRight) == 1);
55     % 总人数=2*总人数-被感染人数+流动人口
56     tableOfArea = ...
57         2 * (tableOfArea == 2) ...
58         - ((tableOfArea == 2) & (sum > 0 | (rand(n
59             , n) < getInfected))) ...
60         + 2 * ((tableOfArea == 0) & rand(n, n) <
61             peopleInGrowth);
62     tableOfArea((tableOfArea == 2) & (rand(n, n) <
63         LongDistanceRun)) = 0;

```

```

58         tableOfArea((tableOfArea == 1) & (rand(n, n) <
           ShortDistanceRun)) = 0;
59     a = find(tableOfArea == 2);
60     b = find(tableOfArea == 1);
61     c = find(tableOfArea == 0);
62     aa = length(a);
63     bb = length(b);
64     cc = length(c);
65     removedpeople(i) = cc;
66     deathPeople(i) = aa;
67     infectedPeople(i) = bb * 30;
68     % 575*575*3 double
69     set(imh, 'cdata', cat(3, (tableOfArea == 1), (
           tableOfArea == 2), zeros(n)))
70     drawnow
71     figure(2)
72     delete(m)
73     plot(0.05 * deathPeople);
74     hold on
75     plot(infectedPeople);
76     hold on
77     % plot(removedpeople);
78     legend(['新增死亡', num2str(bb), ' (人)'], ['
           新增确诊', num2str(aa), ' (人)']);
79     title(['时间T=', num2str(i), '天']);
80     % m = annotation('textbox', [0.15, 0.8,
           0.1, 0.1], 'LineStyle', '-', 'LineWidth',
           1, 'String', str);
81     hold off
82     %控制速度
83     % pause(0.1)
84     end
85     % 拟合

```

```
86     [data, text] = xlsread('dailyAmerican.xlsx');
87     positive = data(:, 3);
88     negative = data(:, 4);
89     pending = data(:, 5);
90     hospitalizedCurrently = data(:, 6);
91     inIcuCurrently = data(:, 8);
92     recovered = data(:, 12);
93     death = data(:, 14);
94     deathIncrease = data(:, 20);
95     hospitalizedIncrease = data(:, 21);
96     negativeIncrease = data(:, 22);
97     positiveIncrease = data(:, 23);
98     death = data(:, 6);
99     x = [1:138];
100     % hold on;
101     % plot(x, deathIncrease, '*r', x, positiveIncrease
102           , '*g');
103     % xlabel('距离1月22日天数', 'FontSize', 12);
104     % ylabel('人数', 'FontSize', 12)
105     % legend('新增死亡', '新增确诊', '拟合死亡', '拟合
106           确诊');
107     % title('图A');
```