



## BDSC CAAI 2021 社会计算大赛

作者单位： 杭州师范大学、浙江大学

作者姓名： 1. 王逸伟

2. 谢明

3. 谢晓雯

4. 王志鹏

5. 王敏

指导教师： 詹秀秀、张子柯

指导教师所在单位： 杭州师范大学、浙江大学



## BDSC CAAI 2021 社会计算大赛

### 题目：新冠疫情背景下北京冬奥会的风险评估：分析、模拟和预测

#### 摘要

作为重大突发卫生公共事件，新冠肺炎疫情给包括东京奥运会在内的大型体育赛事带来了巨大的影响。本文通过对东京奥运会疫情与舆情的分析，发现新冠疫情和东京奥运的舆情之间存在较强的时滞相关性，且总体的社会舆论倾向于消极。此外，我们用长短期记忆网络(LSTM)对北京冬奥会的所有参赛国进行了冬奥会赛前疫情预测。进一步，我们提出改良的SEIR模型，对冬奥会期间奥运村中的病毒传播情况进行了模拟。结果表明，限制活动范围、缩短检测周期可有效地避免封闭区域内疫情的爆发，缩短响应时间也能极大地控制疫情规模。基于以上的数据及模型分析，我们对冬奥会期间疫情和舆情所包含的潜在风险进行了评估，并给出相应的决策参考。

关键词：新冠；奥运会；时序预测；基本再生数；SEIR模型

# 目录

摘要.....	I
目录.....	II
一、研究背景.....	1
二、问题描述.....	1
三、数据集描述与分析.....	2
（一）数据集描述.....	2
（二）相关性分析.....	5
（三）情感分析.....	8
四、基于 LSTM 模型的冬奥会赛前疫情预测.....	9
（一）模型建立.....	9
（二）模型训练.....	10
（三）训练结果与分析.....	12
五、基本再生数估计.....	13
六、基于改良 SEIR 模型的冬奥会赛中疫情预测.....	15
（一）模型建立.....	15
（二）传播结果与分析.....	16
（三）基本再生数和有效再生数的估计.....	20
七、风险分析.....	21
八、总结.....	22
九、参考文献.....	26
十、附录.....	26
附录一：各参赛国疫情相关数据.....	26
附录二：东京奥运会舆情相关数据.....	26
附录三：各国基本再生数估计.....	30
附录四：LSTM 感染人数预测结果.....	33

## 一、研究背景

2019 年末至今,新冠肺炎疫情席卷全球,截至 2021 年 7 月 29 日已有超 1.9 亿人确诊,超 416 万人因此丧生。新冠肺炎疫情扰乱了社会的正常运转,对人们生活造成持久的影响,对全球经济造成巨大冲击以及伤害。具有典型聚集特征的体育赛事,不可避免地受到波及 [1-4]。全球肆虐的新冠肺炎疫情使得世界体育不得不按下“暂停键”:东京奥运会、欧洲杯延期,各国职业联赛“停摆”,各类“大满贯”级别的品牌赛事取消,甚至人们日常的跑步健身活动也因“居家令”、“禁足令”等无法进行。

不断变化的外部环境会给体育发展带来变动不居的多重影响,规约和牵引着体育赛事的走向。面对未知的全球疫情发展情况,梳理与剖析“新冠疫情”给体育带来的影响与冲击,以尽早对可能产生的风险问题采取应对措施,最大程度地减小疫情蔓延对体育赛事产业的不利影响,把握特殊背景下的体育赛事发展机遇,具有重要的现实意义。

作为重大突发公共卫生事件,新冠肺炎疫情对大型体育赛事的冲击是双重的。一方面是疫情冲击,由此造成赛事主办地区以及赛事组织、赞助商等多方的经济损失,以及对运动员身体、心理的极大挑战 [5,6];另一方面是舆情冲击,新冠肺炎疫情的全球大流行增加了体育赛事进程的不确定性,同时也加剧了世界格局的不确定性,这使得舆论环境变得更为错综复杂 [7,8]。而舆情冲击带来的影响或者伤害往往更大。

因新冠肺炎疫情延期一年的东京奥运会已于 7 月 23 日正式开幕。与此同时,担心和争论也从未中断。层出不穷的变异病毒在全球多个国家蔓延 [9]、21 万民众签名反对举办奥运会 [10]、东京的新冠感染情况仍在持续扩大 [11]、部分赛事将采取无观众的空场方式举行 [12]、奥运村出现运动员确诊新冠阳性等等成为了引爆 2020 东京奥运会舆情的一个又一个节点 [13]。

## 二、问题描述

1994 年开始,国际奥委会保持着每两年举办一届奥运会的传统。2022 年北京冬奥会将在东京奥运会举办半年后的 2022 年 2 月 4 日开幕,东京奥运会和北京冬奥会将是二十多年来间隔最近的两届奥运会。毫无疑问,两届国际性体育赛事的间隔时间如此之短,加之国际疫情形势仍不容乐观,北京冬奥会面临着同样严峻的挑战,故着重应对新冠肺炎疫情以及相应的舆情是确保赛事顺利举办的重要一环。

本研究拟通过全球疫情传播数据和东京奥运会的舆情分析,预测北京冬季奥运会的舆情走向,对潜在风险给出决策参考。本研究主要分为七个部分:第一部分给出研究背景,介绍了新冠疫情作为重大突发公共卫生事件给大型体育赛事带来的安全以及舆论挑战;第二部分描述了本研究旨在解决的主要问题,以及本研究的研究框架与思路;第三部分旨在深入研究新冠疫情和东京奥运相关舆情之间的关系,我们首先描述本研究所用的数据,然后分析了全球疫情与东京奥运相关舆情间的时滞相关性,以及疫情背景下东京奥运舆情的情感偏向;第四部分中,我们使用了长短期记忆网络(LSTM)建立机器学习模型,对北京冬奥会的赛前疫情进行预测,评估赛前疫情爆发的风险,为赛前舆情风险提供参考;第五部分中,我们使用疫情数据估计各国在新冠疫情初期的基本再生数;第六部分中,我们对传统 SEIR 模型进行了改进并提出了改良的 SEIR 模型,来模拟奥运村中的疾病传播情况。其中,我们着重研究了隔离强度、检测周期和反应时间三种防疫手段的防疫效果,并对改良的 SEIR 模型进行理论分析,分析结果能更好地指导疫情的防控,从而预先引导积极的社会舆论导向,同时防止消极舆论的爆发;最后一部分中,我们基于舆情的滞后性,以及前两部分的疫情预测结果及分析,对即将来临的北京冬奥会所潜在的疫情风险和舆论风险给出相应的决策参考与总结。

### 三、数据集描述与分析

#### (一) 数据集描述

##### 1. 历届奥运会数据

我们从 Kaggle 开源竞赛网站收集了近 120 年共计 30 届夏季、冬季奥运会参赛国与参赛运动员的数据<sup>1</sup>,最后保留的数据包括运动员编号、所属国家/地区、奥运会届次、参赛年份及赛季(夏季/冬季)。

##### 2. 疫情数据

所用数据集取自约翰霍普金斯大学的系统科学与工程中心 (Center for Systems Science and Engineering, CSSE)<sup>2</sup>,原始数据集收集了从 2020 年 2 月 15 日至 2021 年 6 月 10 日共 229 个国家的完整疫情数据。经处理后只保留每日新增病例数(daily new infections)、每百万新增病例

<sup>1</sup> Rgriffin. (n.d.). 120 Years of Olympic History: Athletes and Results | Kaggle. Retrieved July 29, 2021, from <https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

<sup>2</sup> Our world in data. (n.d.). GitHub - Owid/Covid-19-Data: Data on COVID-19 (Coronavirus) Cases, Deaths, Hospitalizations, Tests • All Countries • Updated Daily by Our World in Data. Retrieved July 29, 2021, from <https://github.com/owid/covid-19-data>

数（new infected cases per million）、总病例数（total infected cases）三个特征。以日本为例，图 1 和图 2 展示了其日新增确诊病例数、每日累计病例数的数据走向（所有参赛国的疫情相关数据见附录一），可以看出每日新增病例数仍持高不下，疫情仍在蔓延，并且世界各国基本都出现了两次新冠疫情爆发的高峰。

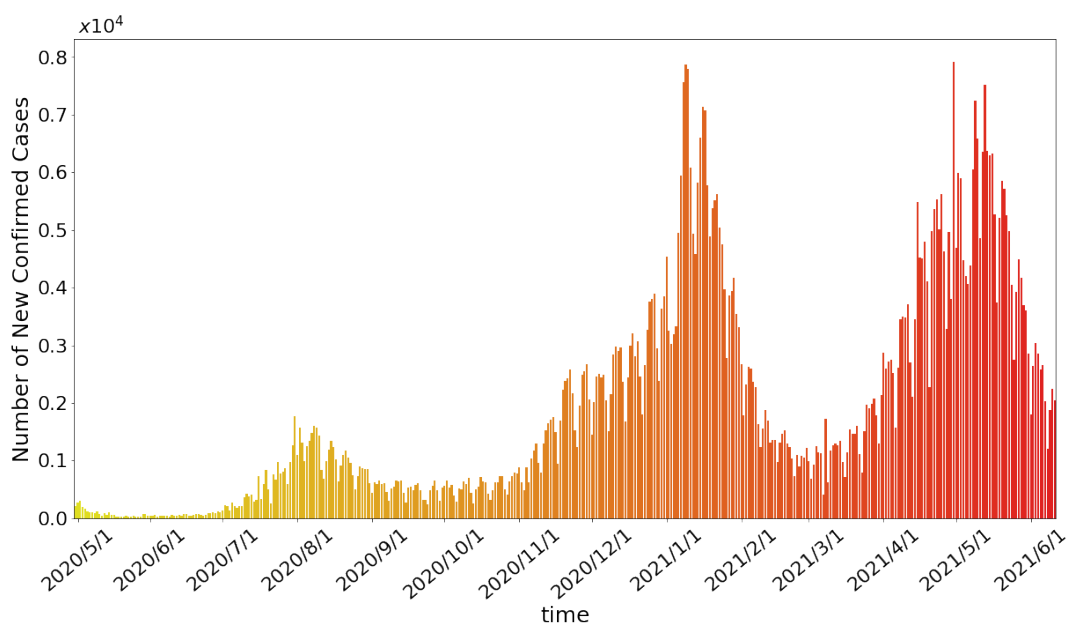


图 1 各参赛国（图为日本）每日新增病例数据

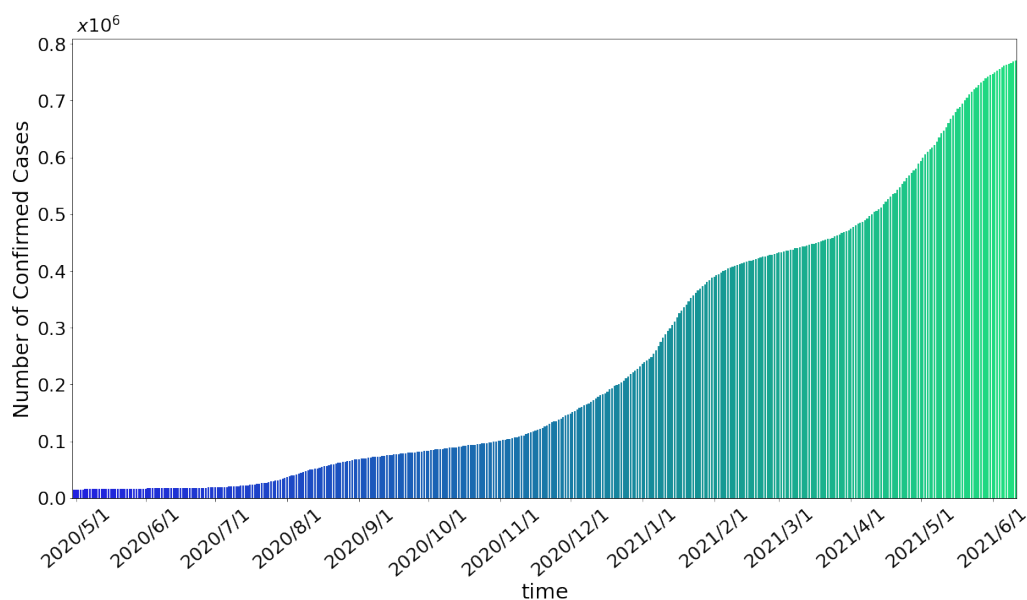


图 2 各参赛国（图为日本）的每日累计病例数据

### 3. 舆情数据

我们以“东京奥运会”为关键词，通过网络爬虫，累计爬取了 47430 条与东京奥运会相关

的新闻报道（舆情的相关数据见附录二）。为了分析国际、中国国内和日本国内针对东京奥运会的舆情走向与新冠疫情的相关性，我们爬取了语言为英语、中文和日语的主流媒体报道。图 3 给出了不同语言来源的新闻报道占比，其中英语、日语和中文的舆情数据占比分别为：42.86%、49.14%和 8%；图 4 给出了不同媒体来源的新闻报道占比，其中的主要来源为路透社（22.6%）、纽约时报（2.49%）、朝日新闻（32.24%）和每日新闻（16.2%）。需要说明的是，有 18.64% 的新闻报道来自不同的非主流媒体，每个媒体的平均报道数较少，故这里合并表示为“其他”。

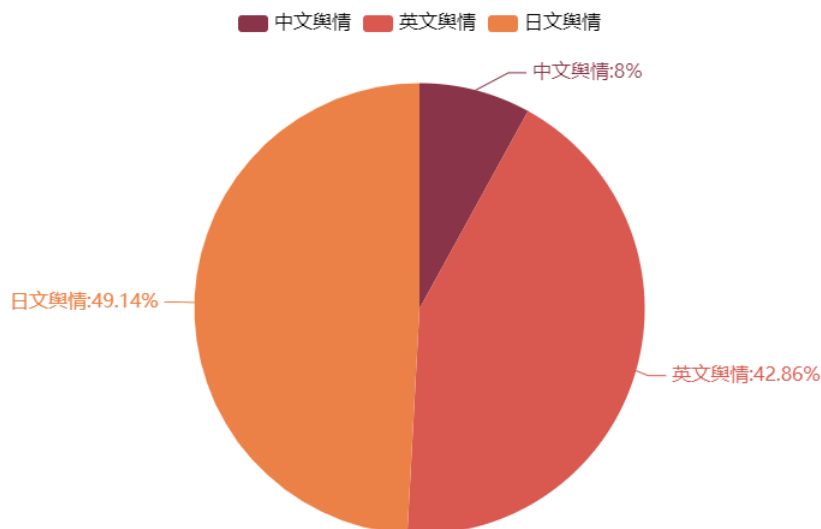


图 3 不同语言的新闻占比

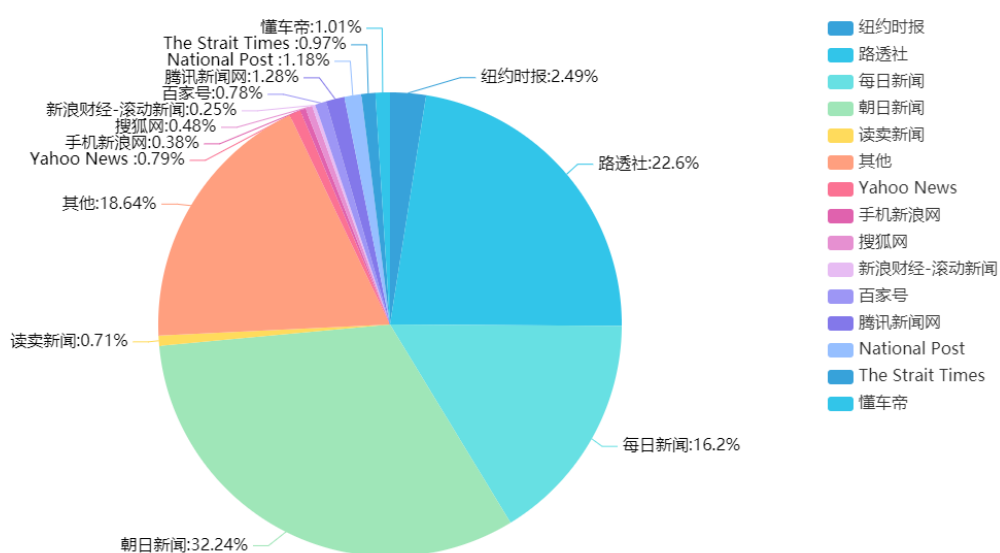


图 4 新闻来源占比

考虑到工具的可获得性与媒体的权威性，我们这里仅对英文新闻报道进行处理和清洗，并

对其做了词云统计（图5），除去与关键词直接相关的单词（如"tokyo", "olympics"等）后，词频前十的词有（以下按照词频降序排列）："coronavirus", "japan", "postpone", "athlete", "pandemic", "health", "vaccine", "public", "virus", "international"。可以看出，媒体对东京奥运会关注的焦点多集中在疫情上。

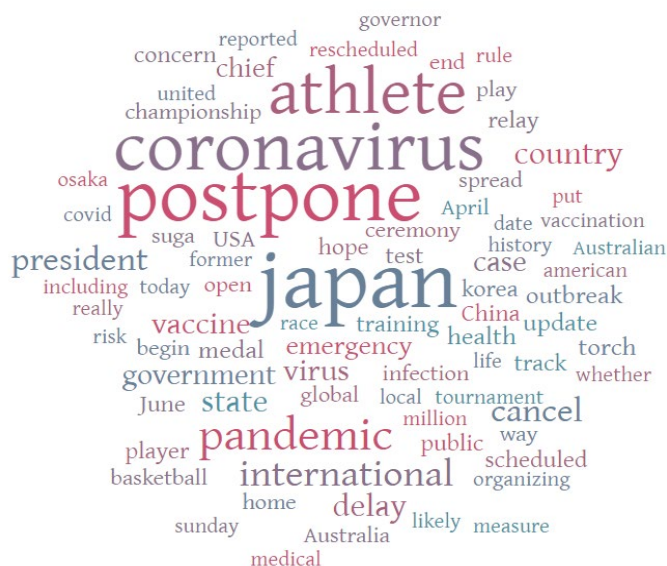


图 5 舆情词云统计

## （二）相关性分析

为研究关于东京奥运会的舆情与疫情之间的关系，我们从国际、中国国内、日本国内三个层面探究了疫情与舆情之间的相关关系。图 6 显示了 2020 年 2 月 15 日至 2021 年 6 月 10 日期间每日的全球新增病例数和东京奥运相关新闻报道数量，其中红色数据表示世界范围内每日关于东京奥运相关的新闻报道数量，蓝色数据表示全球每日疫情的新增病例数。结果表明，全球疫情在 2020 年末和 2021 年 4 月左右达到高峰。东京奥运的社会舆论三次达到高峰。从 2020 年 2 月 15 日关于东京奥运会的第一篇报道开始，就有相当多关于东京奥运会的报道集中在讨论是否会延迟举办，在 2020 年的 2 月 26 日关于是否会延迟的讨论达到了第一次舆论高峰。2020 年 3 月 20 日国际奥委会正式发文称会延迟举办东京奥运会，引来了关于东京奥运会讨论的最高峰。第三次小高峰出现在国际奥委会决定东京奥运会举办时间，即 3 月 24 日。



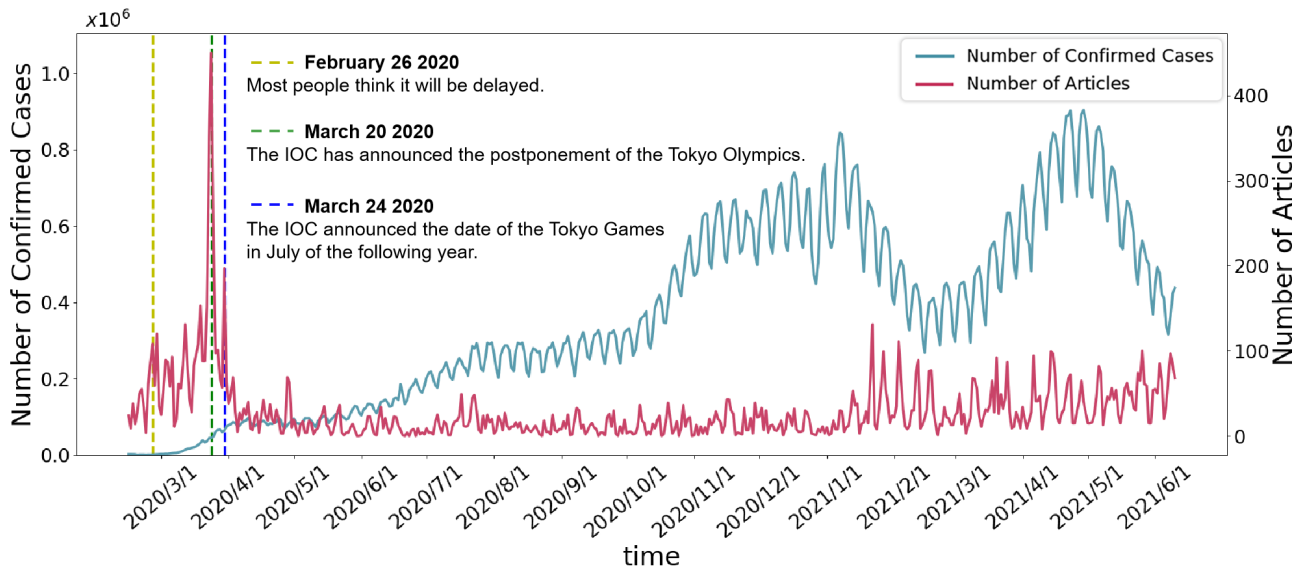


图 6 每日全球新增病例数与东京奥运相关报道数

为了进一步研究在舆论高峰时段，社会舆论的关注焦点，我们对这三个时间段的新闻数据进行了词频统计，最后发现，2020 年 2 月的新闻报道中，"coronavirus", "postpone" 出现频率较高，可见在疫情爆发背景下，社会大众对东京奥运会能否顺利召开高度关注；2020 年 3 月 20 日，国际奥委会鉴于疫情形势，正式宣布推迟举办东京奥运会，引发了又一波舆论高峰；2020 年 3 月 24 日，国际奥委会宣布了东京奥运会推迟举办的具体日期，这一天的舆情词频统计中，"postpone" 出现词频较高（图 7）。

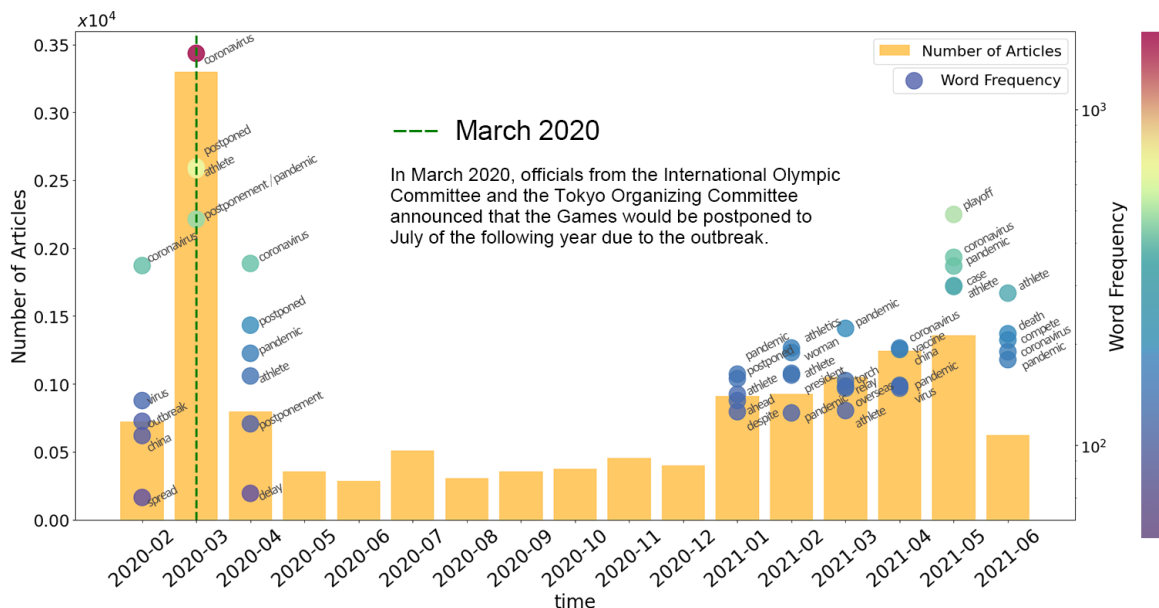


图 7 世界舆情与东京奥运会相关词频的演化图，柱状图为每月舆情总数，图中圆圈为每个月出现的关键词数量，颜色由紫色转红表示该词出现的次数越多。

图 7 中呈现了舆情讨论较多月份的词频统计，从中可以看出，在世界疫情传播的高峰期，有关东京奥运会的舆论焦点多集中在疫情背景下东京奥运会的推迟问题上。

基于以上的分析，我们进一步量化世界疫情和世界舆情之间的皮尔逊相关性，其值为-0.08，即不存在显著线性相关性。疫情的出现引起了关于东京奥运会等赛事的一系列举办以及日程的变化，舆情的爆发相对于疫情的爆发存在一定的滞后性 [14]，因此我们进一步探索两者之间的滞后相关性。我们将疫情的数据向后延迟不同的天数，再计算了疫情与舆情之间的皮尔逊相关性，结果如图 8 所示。图 8 显示，当我们考虑滞后性时，疫情与舆情之间的相关性会随着滞后天数的增加而增大，且最高值能达到 0.54。

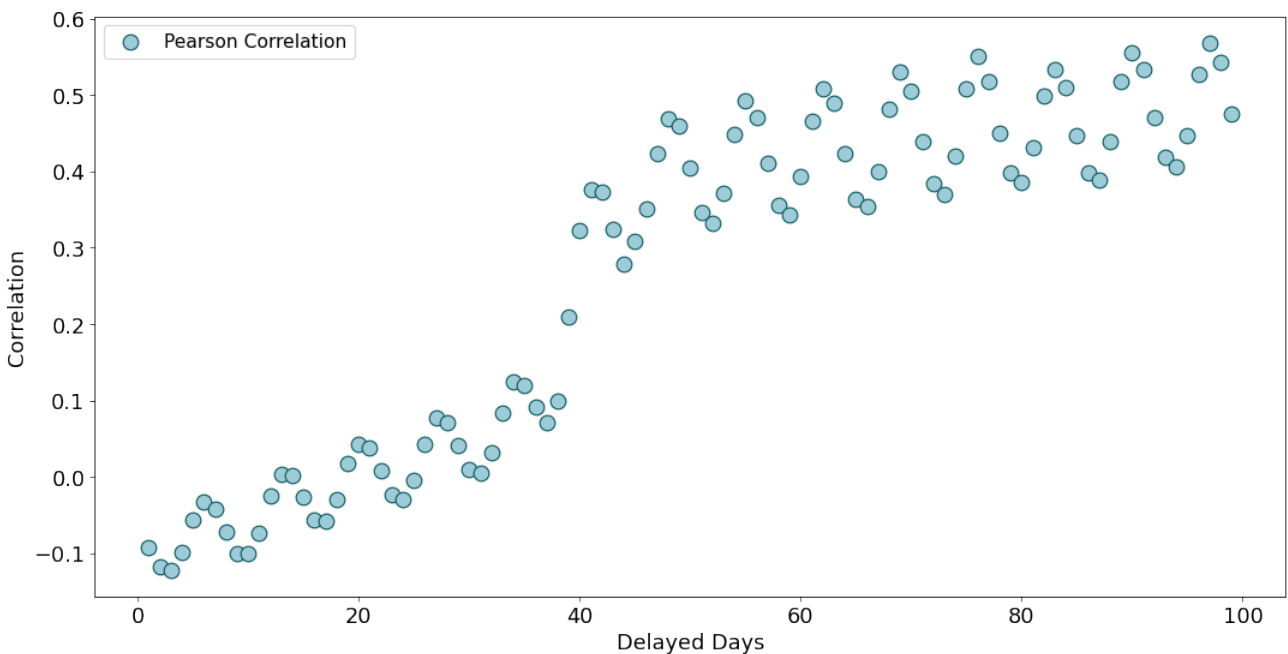


图 8 延迟不同天数下疫情与舆情皮尔逊相关系数的变化

疫情与奥运的舆情是多维度紧密相连的。因此，我们不仅分析了世界范围内的舆情和世界疫情的相关程度，还继续分析了中国相关舆情与国内疫情、中国相关舆情与日本疫情、世界相关舆情与日本疫情、日本相关舆情与世界疫情的关系（如表 1 所示）。表 1 第一行是中国东京奥运日报道数和中国日新增病例数、日本日新增病例的相关性系数，中国针对东京奥运的舆情与日本的疫情发展有更高的线性相关性，可见舆论普遍对赛事举办地的疫情情况持更高的关注度。第二行和第三行中，我们计算了无延迟（黑色）和有延迟（红色）疫情数据两种情况下，疫情和舆情之间的相关性系数，相关性系数后的括号内为相应的  $p$  值。可以看出，疫情和舆情之间的滞后相关性较高（ $>0.5$ ）。

表 1 新增病例与东京奥运相关报道数的相关性系数

$\begin{matrix} \text{\#Infections} \\ \text{\#Articles} \end{matrix}$	Chinese	Japanese	World's
Chinese	0.23 (2.72e-19)	0.39 (1.05e-73)	—
Japanese	—	—	0.38 (2.26e-06) <b>0.65</b> (1.06e-05)
World's	—	0.12 (1.34e-03) <b>0.50</b> (3.05e-03)	-0.08 (2.26e-06) <b>0.54</b> (1.06e-05)

### (三) 情感分析

我们以 2020 年 1 月 10 日至 2021 年 7 月 1 日共 12798 条英文新闻数据为数据来源，按日统计了新闻数据的数量，如图 9 红线所示。为了进一步研究舆论对东京奥运的情感偏向情况，我们使用 SnowNLP 进行舆情数据的情感分析，将新闻数据的标题和摘要作为输入，输出一个 0-1 之间的情感值，越接近 1 表示新闻越积极，越接近 0 表示新闻越消极，同样按日进行计算，将每日新闻计算得到的平均值作为最终的每日新闻情感分数，如图 9 绿线所示。由图 9 我们可以看出，每日的新闻情感值基本都是在 0.6 以下，其平均值为 0.065，这说明关于东京奥运会的社会舆论偏向于消极。

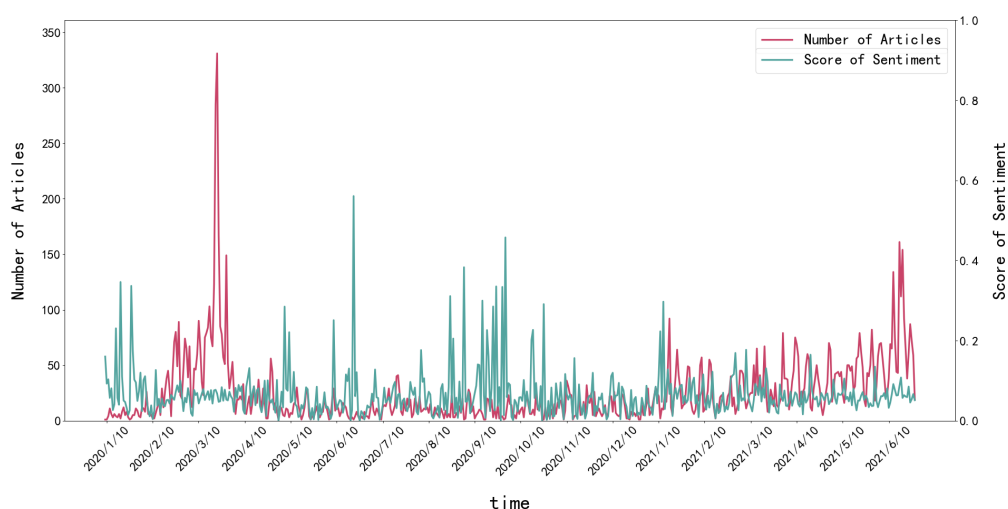


图 9 新闻频率与情感值

#### 四、基于 LSTM 模型的冬奥会赛前疫情预测

社会舆论是否支持奥运这样的国际赛事在疫情背景下召开，很大程度上会参考全球疫情水平，尤其是各参赛国的疫情是否得到有效控制上。因此，在这一部分中，我们基于 LSTM 建立了冬奥会赛前各参赛国疫情的预测模型，以针对不同的疫情控制水平采用不同的舆论引导策略。

##### （一）模型建立

基于疫情与体育赛事之间的相关性，预测疫情不仅能帮助决策者遏制疫情的传播，更能为应对疫情传播带来的舆论挑战提供更多参考。在此部分，我们选用了序列生成模型，以更好地从疫情数据中提取特征并预测未来多天的疫情走向 [15]。序列生成模型，亦称编码解码器模型，该模型最早被应用于 NLP 领域 [16]。给定一个时间序列或单词序列  $X = \{X_1, X_2, \dots, X_n\}$ ，其中  $X_i \in R^a$ ，编码器将  $X$  编码为隐藏特征向量  $Z$ ，其中  $Z \in R^b$ ；之后该向量被输入解码端解码，解码端会据此输出一个序列  $Y = \{Y_1, Y_2, \dots, Y_m\}$ ，其中  $Y_j \in R^c$ 。其基本流程见图 10。

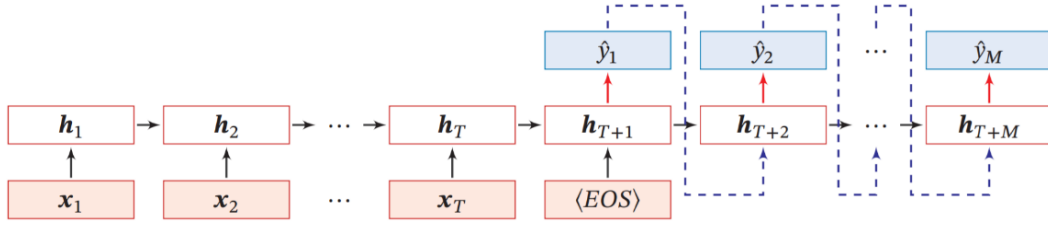


图 10 序列生成模型基本流程 [17]

为保证模型的简洁性和高效性，我们均使用长短期记忆网络（long short-term memory，下称 LSTM）实现编码端和解码端 [17]。作为循环神经网络 RNN 的升级版，LSTM 不仅继承了 RNN 在短时记忆和时序特征上的优点，并且在此基础上引入了内部记忆状态和门控机制。如图 11 所示，每个时刻 LSTM 单元接受上一时刻的输出  $h_{t-1}$  和当前时刻的输入  $x_t$ ，同时存储有上一时刻的内部状态  $c_{t-1}$ ，三者在 LSTM 单元内部受门控的控制，最终生成当前时刻的内部状态  $c_t$  和外部输入  $h_t$ 。对应于控制上一时刻输入、上一时刻输出和当前时刻内部状态的三个门控向量分别为输入门  $i_t$ 、遗忘门  $f_t$  和输出门  $o_t$ ，其生成式如下：

$$\begin{cases} i_t = \text{sigmoid}(W_i \cdot [h_{t-1}, x_t] + b_i) \\ f_t = \text{sigmoid}(W_f \cdot [h_{t-1}, x_t] + b_f) \\ o_t = \text{sigmoid}(W_o \cdot [h_{t-1}, x_t] + b_o) \end{cases} \quad (1)$$

其中  $W_i$ 、 $b_i$  分别为线性变换的权值与偏置， $[*]$  符号表示向量纵向拼接， $\text{sigmoid}(x) = \frac{1}{1+e^{-x}} \in (0,1)$ 。通过门控，模型能够有选择地遗忘无用信息并保留有用信息，从而更有效地提

取时间序列特征。

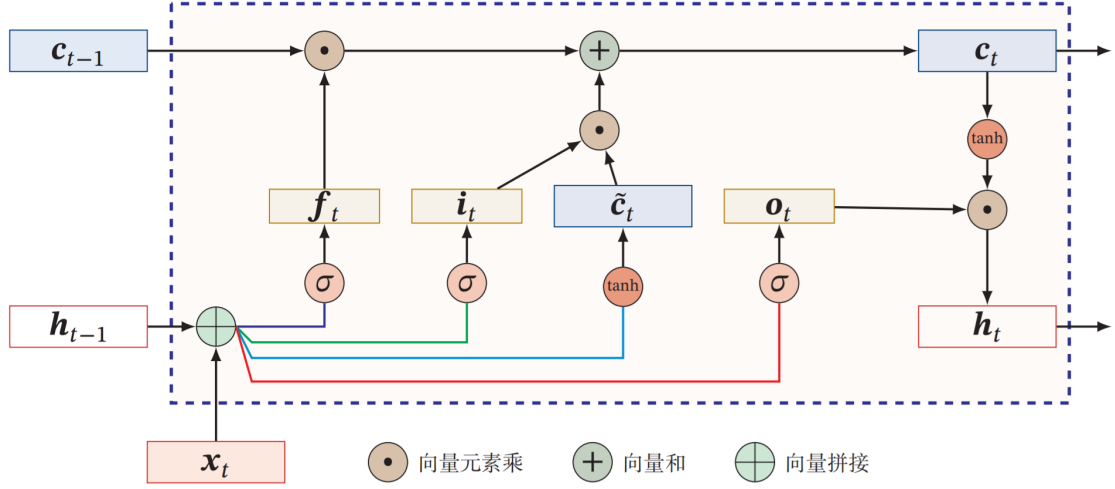


图 11 LSTM 原理

## （二）模型训练

### 1. 数据预处理

我们使用了从 2020 年 2 月 15 日至 2021 年 6 月 10 日共 229 个国家的疫情数据，经处理后我们用每日新增感染人数作为模型的输入特征。对于输入的时序数据，我们首先使用小波变换进行去噪和平滑，步骤如下：

- ① 对含噪信号进行小波分解，选择合适的小波基（此处选用 `sym8`）以及分解尺度，再次进行小波分解，得到一组小波系数；
- ② 对小波分解的各层高频系数进行阈值量化处理，得到小波系数的估计值；
- ③ 对经阈值量化处理的小波系数，进行逆小波变换以重构信号，得到去噪信号。

在第②步中，我们对高频系数使用随等级自适应的量化阈值以及高阶可微的阈值函数进行量化处理 [18]。第  $i$  次小波变换的量化阈值  $T_i$  与阈值函数  $\lambda(x)$  的定义式为：

$$T_i = \frac{\sigma \sqrt{2 \ln N}}{\sqrt[3]{i}} \quad (2)$$

$$\lambda(x) = \begin{cases} x - \frac{x}{(x - T)^2 + 1} & x > T \\ 0 & |x| \leq T \\ x - \frac{x}{(x + T)^2 + 1} & x < -T \end{cases} \quad (3)$$

其中  $N$  为序列长度， $\sigma$  为样本标准差。小波变换的次数是一个经验值，次数过少会导致去噪不足，次数过多又会使得序列过度平滑从而丧失过多信息。基于后面的模型预测表现，我们

对原始序列进行 2 次小波变换，小波去噪效果见图 12。

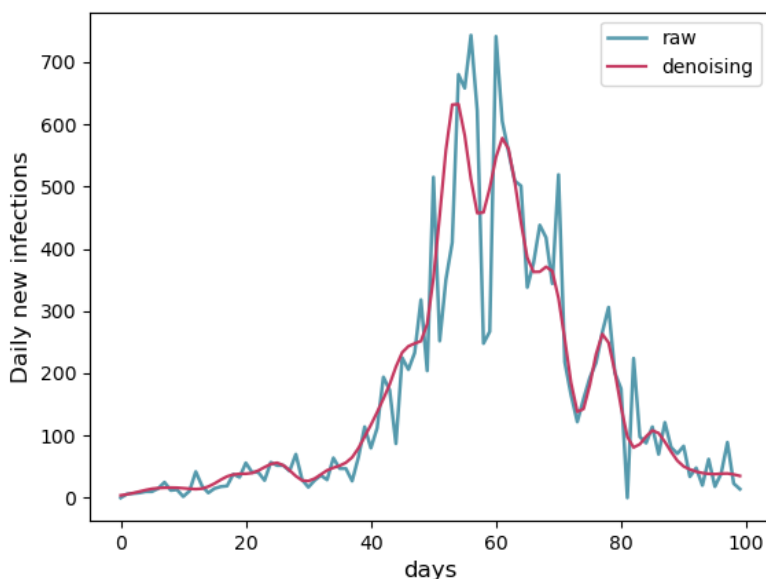


图 12 前 100 天小波去噪效果（本图日本为例），其中蓝色线表示原始数据，红色线表示去噪后的数据

## 2. 数据集分割

我们使用前 90 天的疫情数据预测后 30 天的感染人数走向，原始数据集中收集的每个国家疫情数据从 2020 年 2 月 15 日到 2021 年 6 月 10 日共 482 天，我们将前 70% 作为训练集训练模型，中间 10% 作为验证集寻找最佳参数，最后 20% 数据作为测试集检验模型的预测能力，同时保证未来数据不会出现在历史数据中。使用时间步长为 120 的滑动窗口，将前 90 天数据作为输入，后 30 天数据作为标签构造样本。注意，这里的输入、输出序列长度可任意选取。具体构造方式见图 13。在进行训练时，使用根均方误差作为损失函数，在 python3.6、pytorch-1.7.1GPU 版环境下运行代码，优化算法选用 Adam。

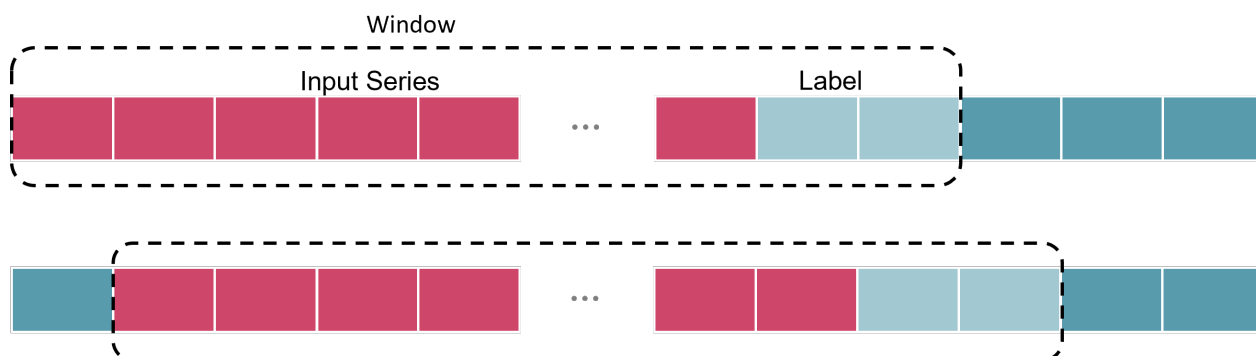


图 13 训练集和测试集构造方法

### （三）训练结果与分析

我们这里使用日本、英国的疫情数据用作模型训练及测试展示，模型训练效果见图 14 和图 15，可以看到 LSTM 能大致预测出感染人数的走向。由于疾病过程中掺杂过多的政治、文化、气候和生物因素，使得疾病感染人数序列具有显著的波动性，即序列的趋势和季节性随时间而变化，使得人工从原始序列中找到内在规律几乎不可能。分析日本在对待疫情方面的政策及最近英国新出现的新冠病毒变种，LSTM 网络可能通过数据捕捉到了日本政府政策的“周期性”及新冠病毒变种的传播规律。同时，随着疫情数据的不断更新，我们的模型也能够从新的数据中进一步学习其内在规律，也即有着“滚动预测”的能力。

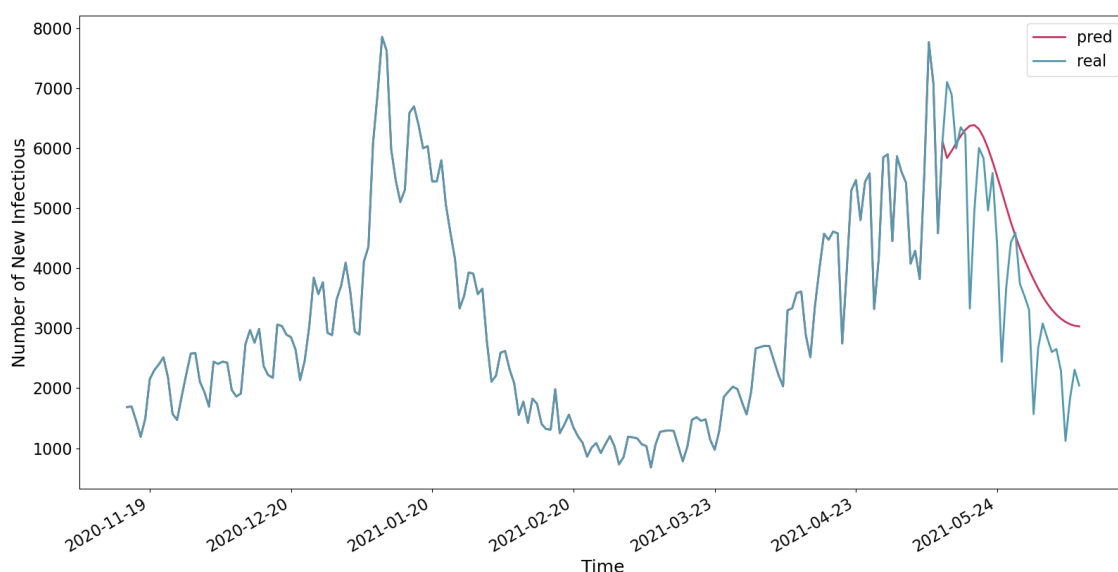


图 14 日本疫情预测结果，其中红色为预测值，蓝色为实际真实值

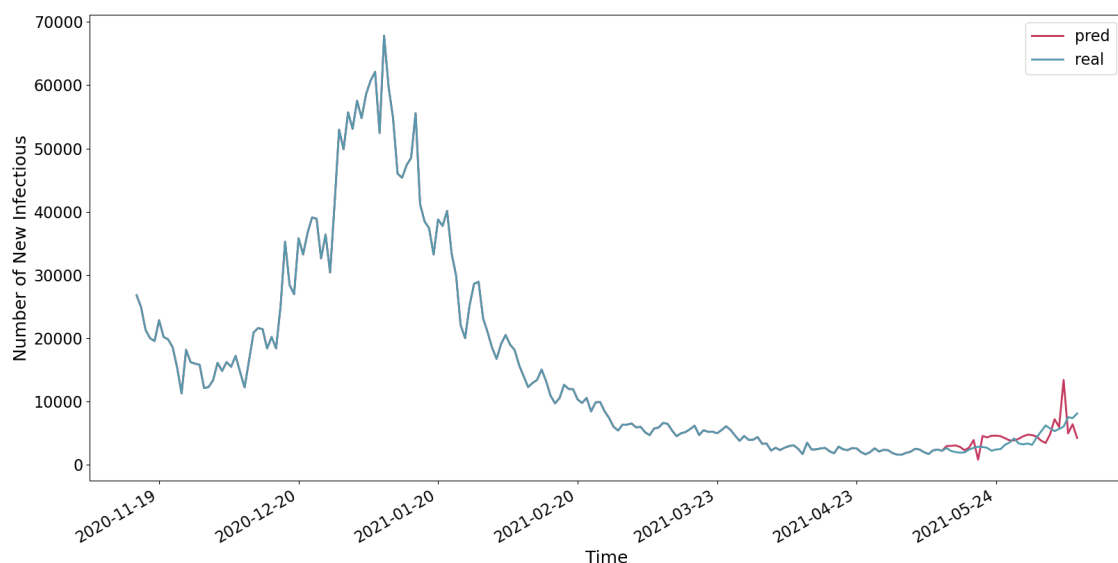


图 15 英国疫情预测结果，其中红色为预测值，蓝色为实际真实值



## 五、基本再生数估计

基本再生数，亦称基本再生率，符号记作 $R_0$ ，是一种流行病学指标，通常解释为患病个体在感染期间所引发的平均感染人数，可用来描述传染病原体的传染性或可传播性。由于传染性是人类行为和病原体生物特征的函数，故在特定场景下的疾病传播，其传染病原体的基本再生数是一个常数。在经典的流行病学模型中（例如 SI、SIS），传染病原体 $R_0$ 的大小决定了其引发的疫情是否会爆发及爆发的规模 [19,20]。由以上可知，基本再生数是一个简洁且有效的流行病衡量指标，它能够帮助我们了解流行病的传播潜力并据此给出合理的防疫措施。

以估计日本新冠疫情的基本再生数为例，所用疫情数据从 2020 年 3 月 15 日开始，结合 4 月 7 日日本政府开始在国内实行紧急状态，我们可以认为 3 月 15 日到 4 月 9 日期间病毒在人群中自由传播，从而可以用经典的 SEIR 模型去建模这 25 天内疫情的传播情况 [21]。该模型将所有人分为 4 种状态：易感者（S）、暴露者（E）、感染者（I）、移除者（R），其中易感者是不携带抗体且不携带病毒的普通人群；暴露者是与感染者接触并携带病毒的群体，暴露者所处的这段时期亦称潜伏期，不具备传播病毒的能力；感染者是出现显性病理症状并具备传播病毒能力的群体，即发病群体；移除者是最后移除病毒的群体，在现实中包括痊愈并产生抗体和恶化并死亡两种情况，无论哪种情况，移除者最后都不可能再次被感染病毒。在平均场假设与总人口不变假设下，描述状态间转移的传播动力学方程为：

$$\begin{cases} \frac{dS}{dt} = -\beta S(t) \frac{I(t)}{N} \\ \frac{dE}{dt} = \beta S(t) \frac{I(t)}{N} - \alpha E(t) \\ \frac{dI}{dt} = \alpha E(t) - \gamma I(t) \\ \frac{dR}{dt} = \gamma I(t) \end{cases} \quad (4)$$

其中 $\beta$ 、 $\alpha$ 、 $\gamma$ 分别为传播率、发病率和移除率，状态转移过程见图 16。

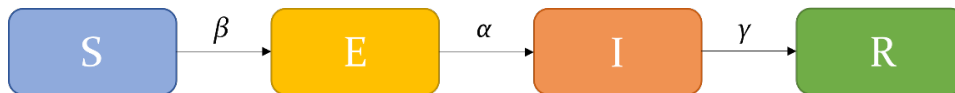


图 16 传统 SEIR 模型

其中 $\beta$ 、 $\alpha$ 、 $\gamma$ 分别为传播率、发病率和移除率。传播率通常被描述为一个接触过感染者的



易感者被感染的概率，发病率定义为潜伏期的倒数，移除率在这里也被称为恢复率，定义为感染期的倒数，这里取 $\alpha = 0.2$  [9]、 $\gamma = 0.1$  [22]， $\beta$ 通过模型拟合实际数据得到，之后我们可根据如下式子计算出基本再生数：

$$R_0 = \frac{\beta}{\gamma} \quad (5)$$

构造优化目标：

$$\beta_s^* = \underset{0 < \beta < 1}{\operatorname{argmin}} J_s(\beta) \quad (6)$$

其中

$$J_s(\beta) = \sum_{t=1}^{25} [\hat{C}_s(t) - C(t) - Y_s(t)]^2 \quad (7)$$

$\beta_s^*$ 是参数 $\beta$ 的估计量，且估计量 $\beta_s^*$ 服从正态分布 [24]； $\hat{C}_s(t)$ 、 $C(t)$ 分别表示累计感染人数的观测值和理论值， $Y_s(t)$ 为观测的累加噪声。构造过程及细节见附录 A 拟合目标的推导。

我们将日本国内前 25 天累计感染人数数据作为 $\hat{C}_s(t)$ 的样本值 $\hat{C}(t)$ 代入式(7)，并使用 Adam 优化器优化式(6)得到参数 $\beta$ 的估计值 $\beta^*$ 。重复计算一万次 $\beta^*$ 可得到图 17 所示的估计量 $\beta_s^*$ 的分布。基于该分布我们可获得 $\beta$ 最终的估计值为 0.28918（95%置信度，0.28914-0.28922）拟合效果见图 18。最终我们将 $\beta$ 的估计值代入式(5)，可得日本新冠病毒传播的基本再生数 $R_0 = 2.8918$ 。

我们以日本为例估算出了该地区新冠病毒传播的基本再生数，其值与 SARS 病毒的基本再生数接近 [25]，表明其传染性处于较高水平。基于上述方法以及其他地区的新冠病毒感染数据，我们估算出了剩余国家或地区的基本再生数（见附录三），在数据真实性的前提下，估计出的基本再生数从侧面反应了初期各国家和地区的社会行为是否有利于疫情的扩散。

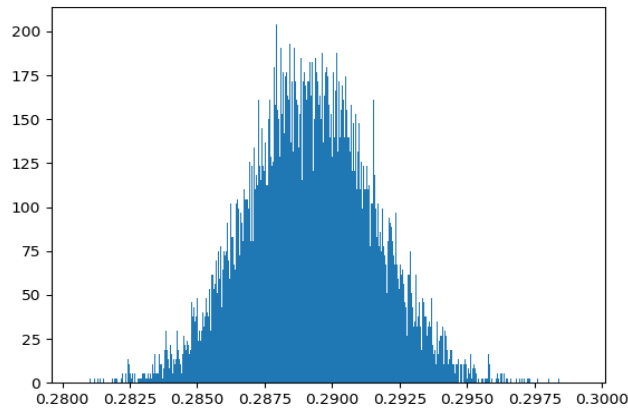


图 17  $\beta_s^*$ 的分布

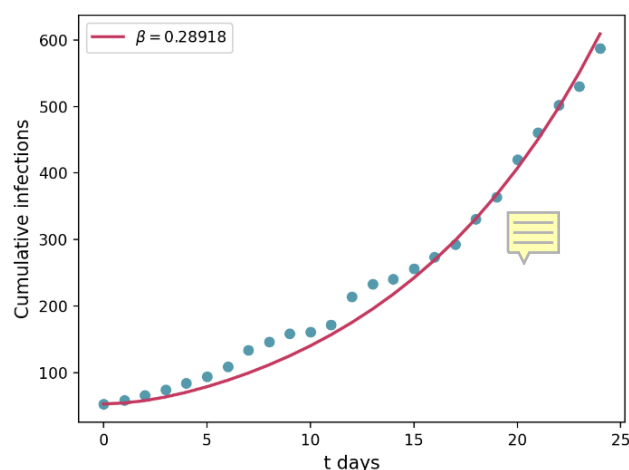


图 18  $\beta$ 取 0.29 的拟合效果

## 六、基于改良 SEIR 模型的冬奥会赛中疫情预测

奥运村的疫情防控措施是否合理、完善，在赛前是影响舆论是否支持奥运会举办的关键因素，在赛中更是一项巨大挑战。奥运村的人员来自世界各地，有不同的文化和信仰，其所在国的疫情控制情况以及疫苗接种普及率等等都有巨大差别。如何建立良好的防控措施，使奥运村内的疫情得到有效控制，也是北京冬奥舆情引导的一个重要方面。在这一部分，我们基于改良的 SEIR 模型，研究了不同隔离强度、检测周期和响应时间水平下的奥运村内疫情传播情况，以供决策者参考。

### （一）模型建立

新冠病毒的传播特性及奥运村的人员接触与原始的 SEIR 模型假设有较大不同。首先，已经发生多起无症状感染者在人群中自由活动从而导致疾病局部爆发的事件，这意味着处于潜伏期的新冠病毒携带者仍然具有传播性。其次，由于防疫意识的提高，人们对于具有显性症状的人的接触意愿大大降低。最后，由于奥运村的安全防护等级较高，出现明显的群体感染症状后，官方会立即采取个体隔离措施并将所有排查到的显状病例送入病房。将上述情况抽象，我们可以列出以下修正的 SEIR 模型：

$$\begin{cases} \frac{dS}{dt} = -\beta_1 r(t) S(t) \frac{E(t)}{N} - \beta_2 r(t) S(t) \frac{I(t)}{N} \\ \frac{dE}{dt} = \beta_1 r(t) S(t) \frac{E(t)}{N} + \beta_2 r(t) S(t) \frac{I(t)}{N} - \alpha E(t) \\ \frac{dI}{dt} = \alpha E(t) - \gamma(t) I(t) \\ \frac{dR}{dt} = \gamma(t) I(t) \end{cases} \quad (8)$$

$S(t)$ 、 $E(t)$ 、 $I(t)$ 分别表示时刻 $t$ 奥运村中易感者数量、潜伏者数量和感染者数量； $R(t)$ 表示时刻 $t$ 奥运村中所有不参与状态转移的人，包括免疫者、死亡者和被发现然后被送入医院的感染者； $\beta_1$ 、 $\beta_2$ 为暴露者传播率与感染者传播率； $t_0$ 为响应时间，表示官方察觉到病毒并加强防疫措施的时间点； $r(t)$ 、 $\gamma(t)$ 分别表示奥运村中每人每日接触的平均人数和卫生检查周期，两者在响应时间前后有不同的值，体现了人们在察觉到疫情后的主动防疫手段； $\alpha$ 表示发病速率。参数的详细介绍见附录 A 表 1

## （二）传播结果与分析

传播实验在 python3.6 环境下进行，我们考虑最差的情况，即所有人员均未获得免疫能力，初始感染者为 1 人。这不仅是应对当前新冠病毒的最差情况，并且能模拟其未知的病毒变株在奥运村的传播情形。我们研究了不同隔离强度、检测周期和响应时间水平下的奥运村内疫情传播情况。

### 1. 隔离强度

隔离强度反映了奥运村限制出行的程度，由接触人数 $r(t)$ 衡量。我们重点关注响应时间后的接触人数 $r_1$ ，它反映了官方在了解到病毒存在后采取的隔离强度，这里我们取 $r_1$ 的值为 5、8 和 12，响应时间 $t_0 = 7$ ，反应期后移除速率 $\gamma_1 = 0.2$ （反应期后检测周期为 5 天），对应奥运村内的感染者、累计感染者及露者数量见图 19 (a-c)。可以看出，疾病爆发程度与隔离程度呈负相关，由于初始状态感染人数较少，在不采取任何加强隔离措施的情况下（ $r_1 = 12$ ），约 68 天疾病爆发达到最高峰；当隔离程度加强时（ $r_1 = 8$ ），最高峰会相应延迟；当隔离强度较大时（ $r_1 = 5$ ），其累计感染人数处于一个极低的状态（120 天时累计感染人数为 200 人）。

通过进一步细化 $r_1$ 的取值（见图 19 (d)），即在 60 天的时间内取 $r_1$ 为 4、5、6，我们发现在 7 天（即反应期 $t_0$ ）后，不同的隔离强度导致了完全不同的感染人数趋势：当反应期后接触人数大于某个阈值时，村中感染人数逐渐增加，疫情呈爆发趋势；当反应期后接触人数小于该

阈值时，感染人数持续走低从而疫情得以控制。

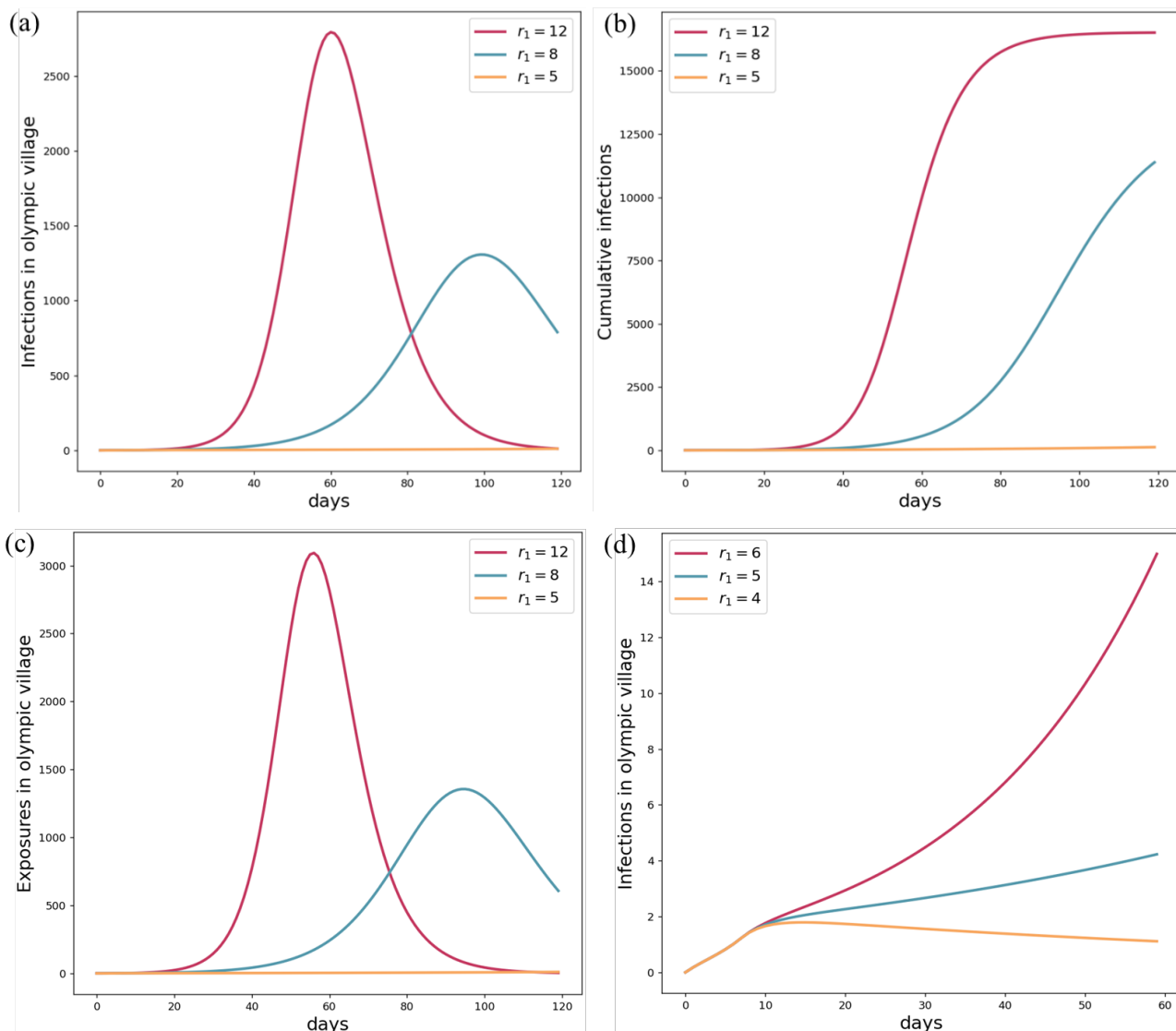


图 19 不同隔离强度下奥运村内群体状态

值得注意的是，奥运村从人员入住到关闭大约 60 天（由于冬奥会期间残奥会人员同居住在奥运村内，故这里的 60 天按冬奥及残奥人员入住至离开的时间算），由于潜伏期的存在，奥运村中暴露者数的变化相比于感染者的变化有 4-5 天的延时，每日暴露者数量大约是当日感染者的 2 倍，且在奥运村闭村时无症状感染者的数量也几乎处于顶峰。奥运会结束后，大量处于潜伏期的人乘坐公共交通回国，途中不会显示出异常状态，在人们几乎无防备的情况下将病毒携带至所到的每一处，或造成新一轮的世界疫情爆发。

## 2. 检测周期

检测周期即奥运村中的人员接受核酸检测的频率，由移除速率 $\gamma(t)$ 衡量。类似于隔离强度，我们重点关注响应时间后的移除速率 $\gamma_1$ ，其值为检测周期的倒数，反映了官方在了解到病毒存

在后对奥运村内人员的检测强度。这里我们取反应后移除速率 $\gamma_1$ 的值为 0.33、0.25、0.2，分别对应反应后检测周期为 3 天、4 天、5 天，反应后隔离人数 $r_1 = 5$  和 8，反应期 $t_0 = 7$ ，对应的群体状态变化见图 20，其中图 20 (a)对应 $r_1 = 5$ ，图 20 (b)对应 $r_1 = 8$ 。

观察图 20 (a)，其各个时刻取值似乎过于小而不符合常理，但仔细分析可知，纵轴表示“当前时刻奥运村中感染者的数量”而非“当前时刻感染者的数量”，且该情形下隔离强度极大 ( $r_1 = 5$ )并伴有周期性的病毒检测，这在现实情况中几乎不可能发生，随着移除速率的增加，当前时刻感染人数的曲线走势明显放缓。类似于隔离强度，移除速率似乎也决定了疫情的走势。结合图 20 (b)可看出，高隔离强度和短检测周期能很好地控制疫情，而低隔离强度和长检测周期则易导致疫情爆发。后面我们将看到，疫情的走向不仅与隔离强度和检测周期有关，还与当前易感者比例相关。

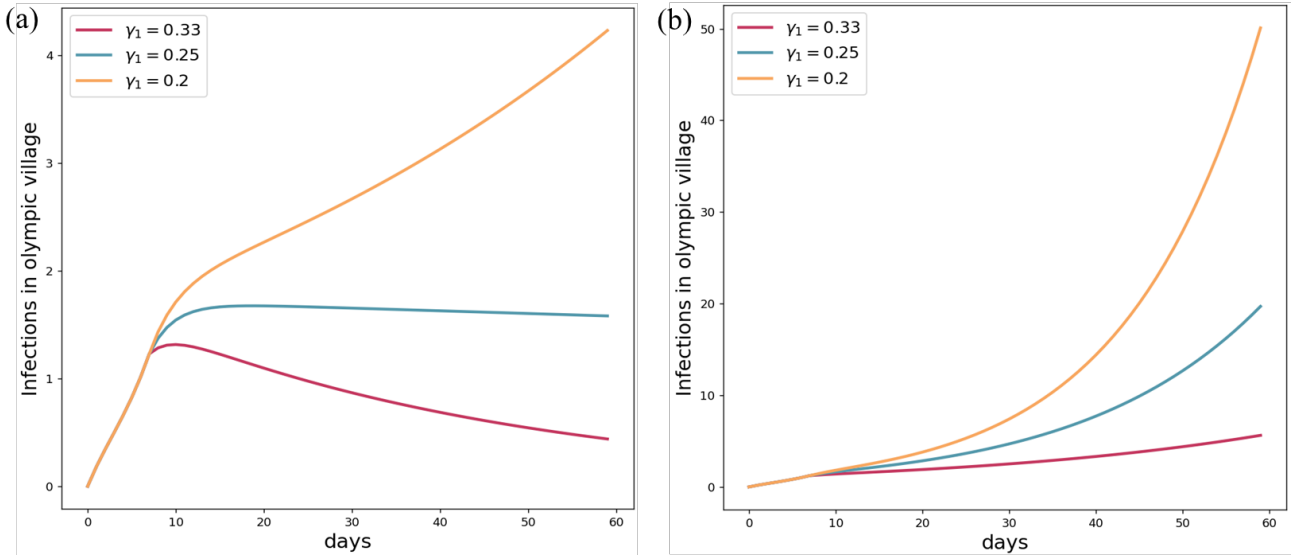


图 20 不同检测周期下奥运村内群体状态

### 3. 响应时间

响应时间反映了官方知晓感染者存在并决策（决定隔离和加强检测）的所用时间，由响应时间 $t_0$ 所衡量。我们假设在反应期后官方会采取两种防疫措施：第一种为极端防疫，官方会采取及其严格的隔离措施并缩短排查病毒的时间间隔，情况可参数化为 $r_1 = 4$ 、 $\gamma_1 = 0.33$ ；第二种为普通防疫，官方此时只会略微限制奥运村中人员的自由活动且稍微缩短排查病毒的时间，此时有 $r_1 = 8$ 、 $\gamma_1 = 0.25$ 。我们取响应时间 $t_0$ 的值为 7、11、15，对应的疫情走势见图 21 ( $r_1 = 4$ )与图 22 ( $r_1 = 8$ )。其中 15 天是新冠病毒的潜伏期+恢复期时长，也是我们改良 SEIR 模型

中 $t_0$ 的最大取值。

观察图 21 (a)与图 22 (a)可知，响应时间的改变并未改变总体疾病趋势。也即，如果隔离强度和检测力度不够，即便加快对疾病的响应时间，也不能阻止疾病的爆发。

然而，加快响应时间虽不能改变疫情的总体趋势，却能大大降低其规模。通过比较图 21(b)与图 22(b)可知，无论是否实施了强有力的措施，加快响应时间均可使疫情的总体规模下降。在隔离强度和检测力度足够的情况下，2 倍的响应时间可使得 60 天后疫情总体规模下降约 6 倍。

根据计算机模拟我们可以看到，更严格的隔离、更频繁的病毒检测能够降低疫情的走势甚至阻断疫情爆发的可能。提高响应时间虽不能改变疫情的走势，却可以大大减小其规模。故防疫期间，在保证有效隔离和高效检测的基础上，决策者需要更快速、更果敢地做出决定。

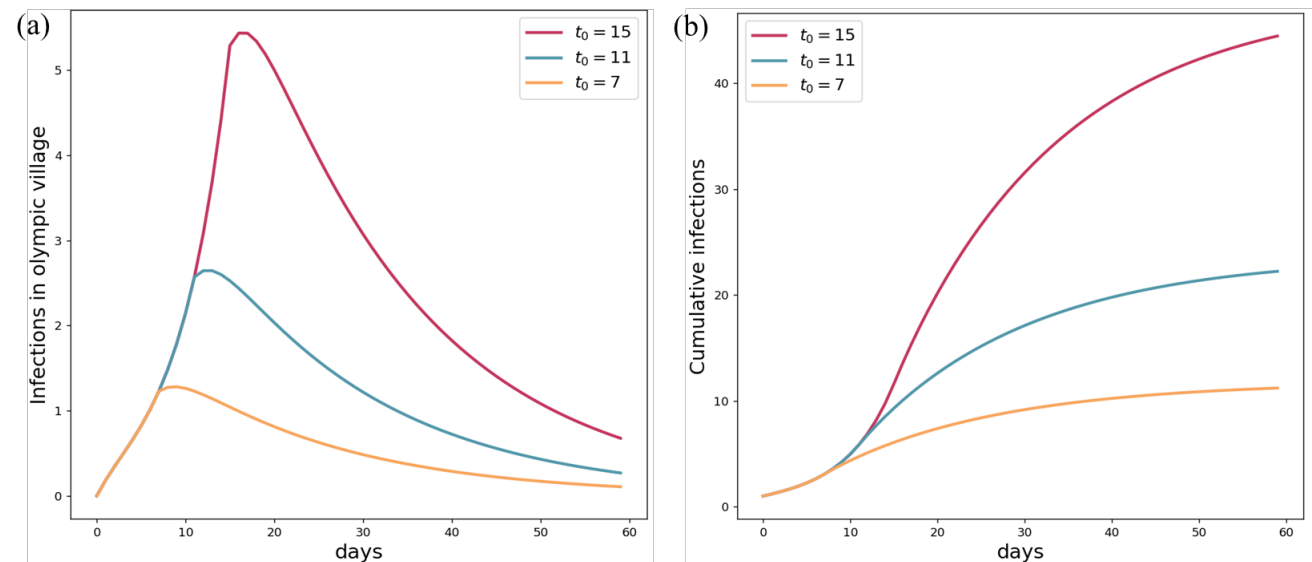


图 21 不同响应时间下奥运村内群体状态 ( $r_1 = 4$ )

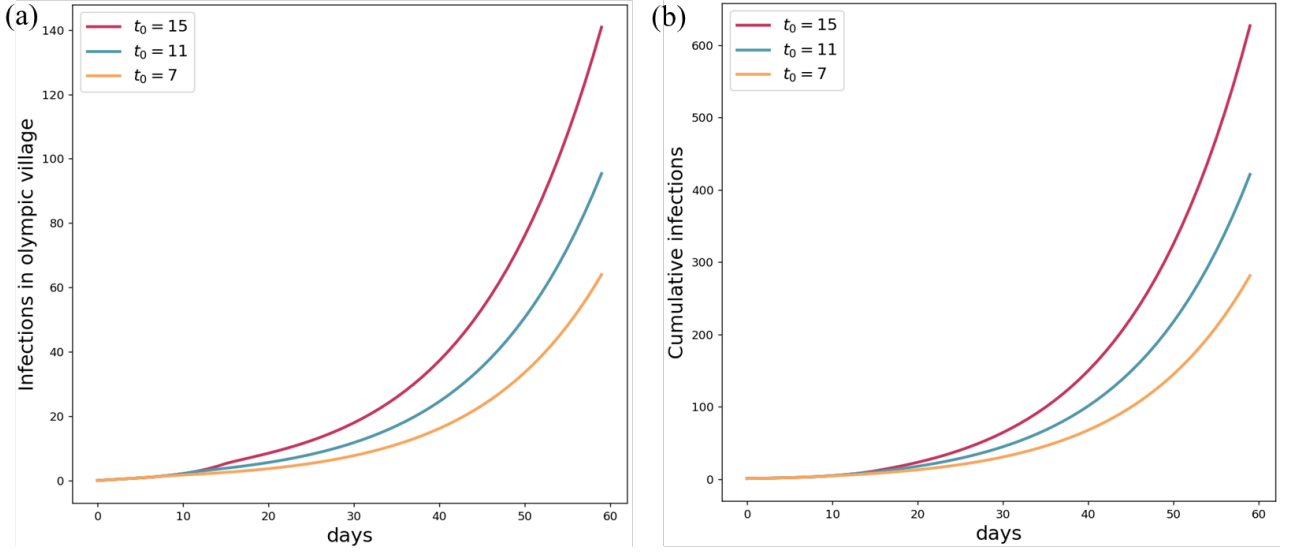


图 22 不同响应时间下奥运村内群体状态 ( $r_1 = 8$ )

### (三) 修正再生数

上述一系列实验结果提醒我们,新冠疫情是否爆发似乎与反应时刻后接触人数及反应时刻后移除速率有关,这促使我们从方程组出发去探究疾病爆发的原因。由附录 A “修正再生数的简要计算”可知,改良 SEIR 模型的基本再生数 $\mathfrak{R}_0$ 和有效再生数 $\mathfrak{R}(t)$ 计算公式为:

$$\mathfrak{R}_0 = \left( \frac{\beta_1}{\alpha} + \frac{\beta_2}{\gamma_1} \right) r_1 \quad (9)$$

$$\mathfrak{R}(t) = \mathfrak{R}_0 s(t) \quad (10)$$

显然 $\mathfrak{R}_0$ 与时刻 $t$ 无关,仅与表示反应时刻后的防疫措施强度参数 $r_1$ 、 $\gamma_1$ 有关。若反应时刻后采取的措施足够有力使得 $\mathfrak{R}_0 < 1$ ,在可预见的时间内,疾病在奥运村中几乎不可能爆发。

与经典 SEIR 模型中类似的是,改良 SEIR 模型中的基本再生数和有效再生数分别描述了疫情是否会爆发以及爆发后疫情接下来的走势,其爆发临界值为 1,也即:当 $\mathfrak{R}_0 < 1$ 时疾病不会爆发,反之则爆发;疾病爆发之后,若 $\mathfrak{R}(t) < 1$ ,则感染者数量减少,反之则增加。与之不同的是,在我们所提出的模型中,两者都与反应时刻后防疫措施相关的参数有关,这能够更好地帮助我们理解在疫情爆发时采取何种措施更有利于控制疫情的传播,从而将舆情导向控制在有利的范围内。所有与 $\mathfrak{R}_0$ 计算相关的防疫措施参数( $\gamma_1$ 、 $r_1$ )均来自于东京奥运会的防疫措施,传播率 $\beta_1$ 、 $\beta_2$ 的估计则基于日本疫情爆发后的前 25 天数据。若在反应后不加强病毒排查,则我们可以根据基本再生数及传播临界值计算出东京奥运村的隔离人数阈值为 $r_1 \approx 4.75$ 。截至截

稿当日，东京奥运会相关人员已累计有 148 人感染，我们的分析指出，如果在奥运村中不采取更强有力的防疫措施，则奥运村中的疫情规模可能会继续扩大。

## 七、风险分析

基于上述长短期记忆模型（LSTM），我们将预测的时间节点延续至北京冬奥会开赛前的 1 月 26 日。通过预测赛前的感染人数数据，以更有效地说明北京冬奥会如期举办的合理性，同时对疫情与舆情的潜在风险做出分析与评估。

在第三部分的分析中，我们得出舆情的走向往往会在疫情爆发后的一段时间内受到显著影响。因此为了引导好北京冬奥会开赛前以及赛事期间舆情的积极走向，防止社会舆论情绪消极化，我们更应预先做好疫情的严格防控。

通过 LSTM 模型的赛前各参赛国感染人数的预测结果见附录四。我们对其风险程度按照赛前的预测感染人数以 0-100、100-1000、>1000 三个范围标准将参赛国划分为低、中、高三个风险等级，详见附录中链接。其中，占低风险等级的有斐济、多米尼克等 16 个国家，占比 17%。中风险和高风险国家分别是 29 个和 50 个，占比为 30%和 53%。分析结果，中国的赛前感染人数预测值为 23 人，处于低风险等级中的较低段，在 95 个参赛国中的风险排名为 89。因此，北京冬奥会在当前的管控力度下如期举办具有合理性。

与此同时，我们也发现各国处于不同的风险等级，美国、巴西、印度等参赛国家处于高风险等级的高段。为了更好判定风险的增势，我们分析了当前（截至 2021 年 7 月 29 日）的风险情况，低、中、高风险等级国家分别为 25，32，38 个（见图 23），占比 26%，34%，40%，与截至 1 月 26 日的风险分析结果对比发现 7 月 29 日以后，多国的疫情风险普遍有上升的趋势，也敦促我们要有更强地防控意识与防控措施。结合上述奥运村中疫情传播的分析结果，北京冬奥会的举办更应注重防控中高风险参赛国的人员输入。从奥运村的疫情模拟结果可知，我们可以通过加强对无症状感染人员、确诊感染人员的隔离程度，提高政策措施的响应时间，合理管控检测周期的参考手段，将其相对应的实际系数降低，从而更好地对奥运会开赛期间的疫情做好防控，进一步预先抑制社会舆情可能存在的消极情绪，引导积极的舆论导向。



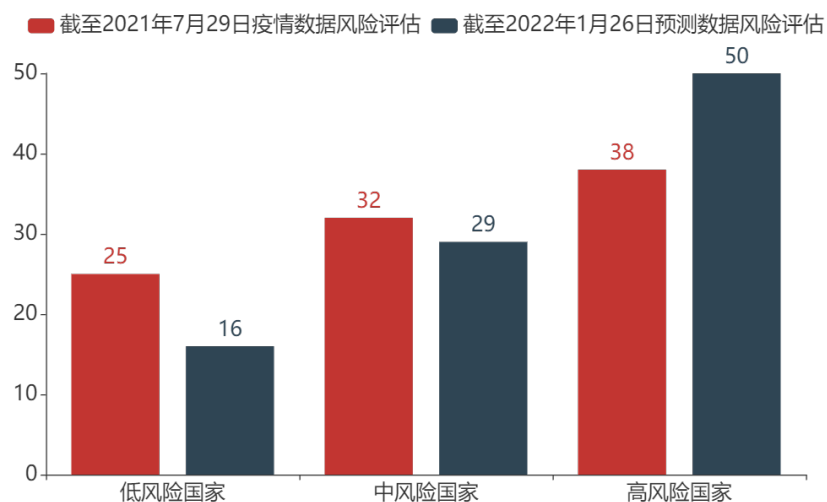


图 23 参赛国风险评估

## 八、总结

本文研究分析了疫情传播与信息扩散两者间的相关性，并对相关舆情进行情感分析，说明了东京奥运会的相关社会舆论普遍偏向消极。在此基础之上，我们进一步分析了世界舆情和世界疫情的时滞相关性。结果表明，全球疫情会在爆发后的一段时间里对日本东京奥运会的相关舆情产生显著影响。同时，文章通过长短期记忆模型（LSTM）预测赛前疫情，针对我国冬奥会能否如期举行的问题提供了理论基础。在我国冬奥会如期举行的前提假设下，本文又进一步对冬奥奥运村中的疫情情况通过改进的 SEIR 模型进行模拟，呈现了隔离强度、检测周期和响应时间三种防控措施在不同程度作用下的疫情传播情况。研究表明，这三种措施对奥运村中疫情的传播和扩散有着有效的抑制作用。最后，我们将赛前预测、赛时模拟和疫情爆发的滞后程度相结合，对疫情和舆情的风险等级进行了评定并做了相关分析。表明了在北京冬奥期间，可以通过加强对参考措施的管理，更加有效地做好大赛时的疫情防控工作，从而更好地引导好未来时滞之后可能发生的舆情导向。

本研究的结果意味着舆情将受到疫情传播程度的滞后影响，不同防控措施等级下对疫情传播的控制也将进而控制舆情的扩散和传播导向，也意味着北京冬奥会若要如期举办，必须严格管控人员隔离、控制隔离周期等，如此才能在疫情流行化的今天，办好奥运会等大型赛事，为世界提供在疫情常态化防控下举办大型赛事并引导积极舆情的中国方案和中国智慧。

## 九、参考文献

- [1] 钟秉枢, 黄志剑, 王凯, 车冰清 & 宋昱. 困境与应对: 聚焦新型冠状病毒肺炎疫情对体育事业的影响. 体育学研究, 2020, 34(2): 9-33.
- [2] 王润斌 & 李慧林. 东京 2020 年奥运会赛事延期交付的多维影响与应对之道. 体育学研究, 2020, 34(3): 7-18.
- [3] 马樱桐. 日本 东京宣布暂停奥运圣火展览活动. 中国会展(中国会议), 2021, (02): 50.
- [4] 张春萍, 杨婧怡, 向佳兴, 周梦渝, 成骆彤 & 陈宇. 新冠肺炎疫情对我国体育赛事利益相关者的影响及应对策略. 北京体育大学学报, 2020, (11): 34-41.
- [5] 叶春明, 赵圣文, 杨秀红 & 刘海韵. 基于机器学习的青少年运动员新冠肺炎疫情应对能力分析与预测. 体育学刊, 2020, (03): 68-73.
- [6] 李佳宸 & 李璐. 新冠疫情时期篮球运动员焦虑成因及调节方法探析. 体育科技文献通报, 2020, (12): 178-180.
- [7] 杨小芳, 李子航 & 贺武华. 从新冠疫情的微博热搜看网络舆情演变及治理. 西南交通大学学报(社会科学版), 2021, (04): 34-40.
- [8] 李坤. 新冠肺炎疫情期间西方的意识形态攻击与中国的应对. 天津师范大学学报(社会科学版), 2021, (04): 70-75.
- [9] 王祎. 世卫警告这种变异病毒即将传遍全球! 有人接种疫苗也感染. 中新网. 2021-7-13, <https://www.chinanews.com/gj/2021/07-13/9518456.shtml>.
- [10] 王珊宁 & 刘凌海. 日本 21 万人签名要求停办东京奥运: 民众感到危险. 海外网. 2021-5-7, <http://news.haiwainet.cn/n/2021/0507/c3541093-32085854.html>.
- [11] 孙璐. 奥运危? 东京一周内新冠感染人数同比增长 146%. 凤凰网. 2021-7-19, <https://sports.ifeng.com/c/8801rwVHBA1>.
- [12] 王丹. 东京都第四次进入紧急状态, 奥运会部分赛事将空场举办, 日本民众感到郁闷…….

环球网. 2021-7-9, <https://3w.huanqiu.com/a/de583b/43rRdCUiV6I>.

[13] 詹婧. 奥运村出现运动员确诊新冠阳性 两名选手同队同项目. 新华网. 2021-7-9, [http://www.xinhuanet.com/2021-07/18/c\\_1127667316.htm](http://www.xinhuanet.com/2021-07/18/c_1127667316.htm).

[14] Zhan X, Zhang K, Ge L, et al. Exploring the effect of social media and spatial characteristics during the COVID-19 pandemic in China. 2021. arXiv preprint arXiv:2107.03766.

[15] Chimmula, V. K. R., & Zhang, L. Time series forecasting of COVID-19 transmission in Canada using LSTM networks. *Chaos, Solitons & Fractals*, 2020, 135: 109864.

[16] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078. 2014.

[17] 邱锡鹏. 神经网络与深度学习. 2021-05-18. <https://nndl.github.io/>.

[18] 陈家益, 战荫伟, 曹会英, 吴兴达, & 李小飞. 连续可微阈值函数与尺度阈值的小波去噪. *电子测量与仪器学报*, 2018, (10): 169-176.

[19] Delamater, P. L., Street, E. J., Leslie, T. F., Yang, Y. T., & Jacobsen, K. H. Complexity of the basic reproduction number ( $R_0$ ). *Emerging infectious diseases*, 2019, 25(1): 1.

[20] Nishiura, H. Mathematical and statistical estimation approaches in epidemiology (pp. 103-121). G. Chowell, J. M. Hyman, L. M. Bettencourt, & C. Castillo-Chavez (Eds.). Dordrecht: Springer Netherlands. 2009.

[21] Bentout, S., Chekroun, A., & Kuniya, T. Parameter estimation and prediction for coronavirus disease outbreak 2019 (COVID-19) in Algeria. *AIMS Public Health*, 2020, 7(2): 306.

[22] Sun, H., Qiu, Y., Yan, H., Huang, Y., Zhu, Y., & Chen, S. X. Tracking and predicting COVID-19 epidemic in China mainland. *medRxiv*. 2020.

[24] Capaldi, A., Behrend, S., Berman, B., Smith, J., Wright, J., & Lloyd, A. L. Parameter estimation and uncertainty quantification for an epidemic model. *Mathematical biosciences and engineering*, 2012, 553.

[25] DataCastle. 周涛：武汉肺炎基本再生数与 **SARS** 接近，确属可防可控的传染病. 腾讯新闻. 2020-1-2, <https://new.qq.com/omn/20200126/20200126A08P9W00.html?pc>

## 十、附录

所有完整数据已上转至 github (<https://github.com/DDMXIE/BDSC2021-Results>)，以下附录提供数据与结果的概览。

### 附录一：各参赛国疫情相关数据

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...
0	AFG	Asia	Afghanistan	2020/2/24	1.0	1.0	NaN	NaN	NaN	NaN	...
1	AFG	Asia	Afghanistan	2020/2/25	1.0	0.0	NaN	NaN	NaN	NaN	...
2	AFG	Asia	Afghanistan	2020/2/26	1.0	0.0	NaN	NaN	NaN	NaN	...
3	AFG	Asia	Afghanistan	2020/2/27	1.0	0.0	NaN	NaN	NaN	NaN	...
4	AFG	Asia	Afghanistan	2020/2/28	1.0	0.0	NaN	NaN	NaN	NaN	...
...	...	...	...	...	...	...	...	...	...	...	...
94852	ZWE	Africa	Zimbabwe	2021/6/6	39189.0	21.0	35.000	1606.0	1.0	1.714	...
94853	ZWE	Africa	Zimbabwe	2021/6/7	39238.0	49.0	39.571	1611.0	5.0	2.429	...
94854	ZWE	Africa	Zimbabwe	2021/6/8	39321.0	83.0	46.143	1617.0	6.0	2.571	...
94855	ZWE	Africa	Zimbabwe	2021/6/9	39432.0	111.0	57.286	1622.0	5.0	3.286	...
94856	ZWE	Africa	Zimbabwe	2021/6/10	39496.0	64.0	57.714	1626.0	4.0	3.143	...

图 24 各参赛国疫情数据概览

### 附录二：东京奥运会舆情相关数据

	news_title	news_content	news_source	news_pubtime	news_url
0	About 10,000 volunteers for the Tokyo Olympics...	The departures were the latest sign of opposit...	New York Times	2021/6/3	<a href="https://www.nytimes.com/2021/06/03/world/asia/...">https://www.nytimes.com/2021/06/03/world/asia/...</a>
1	The Tokyo Olympics, Explained	After a year's delay, the Tokyo Summer Olympic...	New York Times	2021/4/6	<a href="https://www.nytimes.com/article/olympics-tokyo...">https://www.nytimes.com/article/olympics-tokyo...</a>
2	Australia's softball players are among the fir...	The team will be confined to one floor of a ho...	New York Times	2021/6/1	<a href="https://www.nytimes.com/2021/06/01/world/asia/...">https://www.nytimes.com/2021/06/01/world/asia/...</a>
3	Japan extends a state of emergency until one m...	With infections still high, restrictions in ni...	New York Times	2021/5/28	<a href="https://www.nytimes.com/2021/05/28/world/asia/...">https://www.nytimes.com/2021/05/28/world/asia/...</a>
4	Putting Women on More Equal Footing at the Oly...	Almost half of the athletes this year will be ...	New York Times	2021/5/28	<a href="https://www.nytimes.com/2021/05/28/sports/olym...">https://www.nytimes.com/2021/05/28/sports/olym...</a>
...	...	...	...	...	...
13968	Tokyo converts Olympics public-viewing site to...	Tokyo officials said a public-viewing event fo...	Washington Times	2021/6/1	<a href="https://www.washingtontimes.com/news/2021/jun/...">https://www.washingtontimes.com/news/2021/jun/...</a>
13969	Japan's vaccine push ahead of Tokyo Olympics l...	The Olympics risk becoming an incubator for "a...	Denver Post	2021/6/1	<a href="https://www.denverpost.com/2021/06/01/japan-co...">https://www.denverpost.com/2021/06/01/japan-co...</a>
13970	Olympic athletes to sign waiver acknowledging ...	All athletes competing in the Tokyo Olympics m...	Irish Independent	2021/6/1	<a href="https://www.independent.ie/sport/other-sports/...">https://www.independent.ie/sport/other-sports/...</a>
13971	The show must go on	The Australian national handball team arrives ...	Korea Joongang Daily	2021/6/1	<a href="https://koreajoongangdaily.joins.com/2021/06/0...">https://koreajoongangdaily.joins.com/2021/06/0...</a>
13972	Badminton team aim for gold in every event at ...	Korea will contest all five badminton events a...	Korea Joongang Daily	2021/6/1	<a href="https://koreajoongangdaily.joins.com/2021/06/0...">https://koreajoongangdaily.joins.com/2021/06/0...</a>

图 25 英文舆情数据概览

	news_title	news_content	news_source	news_pubtime	news_url
0	東京五輪 一部聖火ランナーの走行市町が決まる / 岐阜	4月4、5日に県内で実施される東京2020オリンピック聖火リレーで、県実行委員会は7日、一部...	毎日新聞	2020/3/8	https://mainichi.jp/articles/20200308/ddl/k21/...
1	東京2020オリンピック聖火リレーの走行時間 / 奈良	◆東京2020オリンピック聖火リレーの走行時間(目安) ①3000【11日】(1) 9:10~...	毎日新聞	2021/4/11	https://mainichi.jp/articles/20210411/ddl/k29/...
2	聖火リレー、警備訓練 盛岡で県警 一連の動作確認 / 岩手	聖火リレーが県内で6月16日~18日に行われるのを前に県警は21日、盛岡市の盛岡競馬場で警備...	毎日新聞	2021/5/22	https://mainichi.jp/articles/20210522/ddl/k03/...
3	書籍で五輪盛り上げる 岡山市立中央図書館展示 / 岡山	2021年に延期された'・'パラリンピックを盛り上げようと、岡山市立中央図書館で書籍の企...	毎日新聞	2020/8/23	https://mainichi.jp/articles/20200823/ddl/k33/...
4	聖火リレー 観覧者募集 出発式・セレブレーション / 滋賀	県は、5月27、28日に県内で実施される'聖火リレーの出発式と聖火の到着を祝うイベント...	毎日新聞	2021/4/22	https://mainichi.jp/articles/20210422/ddl/k25/...
...	...	...	...	...	...
12639	滋賀) 武豊騎手、「しがスपोर्टス大使」に就任	2018年には史上初の通算4千勝を達成するなど、数々の記録を残している。 JRAのレースは、...	asahi_digital	2020/3/22	https://www.asahi.com/articles/ASN3P7D17N3LPTJ...
12640	(しつもん! ドラえもん: 3616) せいか編	東京五輪(とうきょうりん)の聖火(せいか)リレーで、第(だい)1走者(そうしゃ)を務(つ...	asahi_digital	2020/3/21	https://www.asahi.com/articles/DA3S14411069.html
12641	(しつもん! ドラえもん: 3615) せいか編 こたえ	福島県(ふくしまけん) 東京五輪(とうきょうりん)は2011年(ねん)の東日本大震災(ひ...	asahi_digital	2020/3/20	https://www.asahi.com/articles/DA3S14409542.html
12642	バレーのテストマッチは中止	日本バレー協会は16日、4月21~26日に予定していた男女日本代表のテストマッチを中止する...	asahi_digital	2020/3/17	https://www.asahi.com/articles/DA3S14405064.html
12643	鈴木リカルドが初優勝 テコンドー・全日本選手権	テコンドーの全日本選手権は24日、大分県中津市のダイハツ九州アリーナで行われ、男子68キ...	asahi_digital	2020/2/25	https://www.asahi.com/articles/DA3S14378491.html

图 26 日文舆情数据概览

	news_title	news_content	news_source	news_pubtime	news_sentiment
0	强调男女平等原则 石川佳纯成为东京奥运会宣誓候选人	东京奥运会在即，开幕式宣誓代表人选目前仍未完全确定。东京奥运会基于男女平等的原则，正在探讨由...	新传宽频	2021-01-01	0.0
1	朝鲜退出东京奥运会 影响几何?或致朝美关系陷入僵局_中华明法网www.zhmingfa.com	梦见自己当律师给自己辩护朝鲜退出东京奥运会 影响几何?或致朝美关系陷入僵局_中华明法网www...	科技资讯	2021-04-06	0.0
2	日官员称东京奥运举办可能性极大 将探讨观众入场_中华明法网www.zhmingfa.com	苏州打官司最有名的律师日官员称东京奥运举办可能性极大 将探讨观众入场_中华明法网www.z...	国内新闻网	2021-04-13	0.0
3	Shams: 哈登承诺将参加东京奥运, 携手杜兰特并肩作战	中国体育6月22日讯, 据知名记者Shams报道, 篮网球星哈登已经承诺将参加东京奥运会, 届时他...	新传宽频	2021-01-01	0.0
4	哔哩哔哩专栏	随着疫情的原因导致2020东京奥运会延期举办, 同时这也是首届进行延期的奥运会。直到2021年...	哔哩哔哩	2021-04-01	0.0
...	...	...	...	...	...
3791	#东京奥运会赛程场地敲定#从奥运会可能取消到推迟一年举办, 东京奥运会经历了一段让人难以忘记的...	#东京奥运会赛程场地敲定#从奥运会可能取消到推迟一年举办, 东京奥运会经历了一段让人难以忘记的...	懂车帝	2020-07-18	0.0
3792	要推迟了! 中国女足能否拿下最后一张入场券? 入驻2021东京奥运会!	说起中国足球, 读者都默认是中国男足, 不要忘了, 还有一个中国女足在这里。可绝大多数人都选择性...	腾讯新闻网	2020-07-19	0.0
3793	东京奥运刘诗雯冲击首金! 许昕身兼三项没问题, 14天决定5金归属	就在昨日晚间国际奥委会公布了明年东京奥运会的具体赛程, 乒乓球比赛持续时间较长共14天, 东京奥...	懂车帝	2020-07-18	0.0
3794	女排“假球队”有诡计? 中国女排要当心被害, 东京奥运前途凶险	明年的东京奥运会是中国女排队员心目中最大的梦想, 如果能够在东京奥运会上夺得冠军, 中国女排将书...	百家号	2020-07-19	1.0
3795	东京奥运会比赛日程确定 具体什么时候举办?	[#东京奥运会比赛日程和场地已敲定#: 为期17天 共设33个大项339个小项比赛]今天, 东京...	深港在线	2020-07-18	0.0

图 27 中文舆情数据概览

## 附录三：各国基本再生数估计

<b>index</b>	<b>country</b>	<b>left</b>	<b>right</b>	<b>middle</b>	<b>R0</b>
0	Afghanistan	0.38132	0.381862	0.381591	3.81591
1	Albania	0.680402	0.680612	0.680507	6.805068
2	Algeria	0.606688	0.60703	0.606859	6.068592
3	Andorra	0.743392	0.743648	0.74352	7.435203
4	Angola	0.392328	0.392866	0.392597	3.925967
5	Anguilla	0.006544	0.008007	0.007276	0.072756
6	Antigua And Barbuda	0.329182	0.3298	0.329491	3.294912
7	Argentina	0.942553	0.94273	0.942642	9.426416
8	Armenia	0.794916	0.795153	0.795034	7.950344
9	Aruba	0.47472	0.475042	0.474881	4.748812
10	Australia	0.347429	0.347707	0.347568	3.475681
11	Austria	1	1	1	10
12	Azerbaijan	0.555704	0.556081	0.555892	5.558923
13	Bahamas	0.470857	0.471306	0.471082	4.710816
14	Bahrain	0.547489	0.547662	0.547576	5.475757
15	Bangladesh	0.376436	0.376758	0.376597	3.76597
16	Barbados	0.474862	0.475182	0.475022	4.75022
17	Belarus	0.593304	0.593656	0.59348	5.934803
18	Belgium	0.5	0.5	0.5	5
19	Belize	0.364825	0.365399	0.365112	3.65112
20	Benin	0.422278	0.422787	0.422532	4.225324
21	Bermuda	0.422994	0.423351	0.423172	4.231722
22	Bhutan	0.07557	0.080176	0.077873	0.778728
23	Bolivia	0.559641	0.559905	0.559773	5.597728
24	Bosnia And Herzegovina	0.691489	0.691691	0.69159	6.9159
25	Botswana	0.243679	0.244108	0.243893	2.438934
26	Brazil	0.986356	0.986582	0.986469	9.864689
27	British Virgin Islands	0.020844	0.023778	0.022311	0.223112
28	Brunei Darussalam	0.708043	0.708324	0.708184	7.081835
29	Bulgaria	0.754271	0.754453	0.754362	7.54362
30	Burkina Faso	0.827108	0.827331	0.82722	8.272195
31	Burundi	0.16546	0.166123	0.165792	1.657916
32	Cabo Verde	0.27301	0.273702	0.273356	2.733558
33	Cambodia	0.5	0.5	0.5	5
34	Cameroon	0.551765	0.552036	0.551901	5.519009
35	Canada	0.280186	0.280452	0.280319	2.803188
36	Caribbean Netherlands	0.08349	0.084377	0.083933	0.839333
37	Cayman Islands	0.467619	0.468076	0.467848	4.678476
38	Central African Republic	0.243847	0.244606	0.244227	2.442265
39	Chad	0.327453	0.328069	0.327761	3.277608
40	Channel Islands	0.727884	0.728152	0.728018	7.280179

index	country	left	right	middle	R0
41	Chile	1	1	1	10
42	China	0.557458	0.560685	0.559071	5.590712
43	China Hong Kong Sar	0.092429	0.092584	0.092507	0.925067
44	China Macao Sar	3.58E-09	5.00E-09	4.29E-09	4.29E-08
45	Colombia	1	1	1	9.999998
46	Comoros	0.566056	0.566416	0.566236	5.662361
47	Congo	0.500175	0.500602	0.500389	5.003887
48	Costa Rica	0.8291	0.829319	0.829209	8.292094
49	Cote D Ivoire	0.785472	0.78571	0.785591	7.855913
50	Croatia	0.639273	0.639588	0.63943	6.394304
51	Cuba	0.614379	0.614572	0.614476	6.144759
52	Curacao	0.32073	0.321357	0.321043	3.210432
53	Cyprus	0.715892	0.716085	0.715988	7.159885
54	Czech Republic	0.90417	0.90428	0.904225	9.042248
55	Democratic Republic Of The Congo	0.681582	0.681874	0.681728	6.817284
56	Denmark	1	1	1	10
57	Djibouti	0.714083	0.714351	0.714217	7.142167
58	Dominica	0.378585	0.379148	0.378867	3.788668
59	Dominican Republic	0.822913	0.823129	0.823021	8.230213
60	Ecuador	1	1	1	10
61	Egypt	0.507321	0.507733	0.507527	5.07527
62	El Salvador	0.672695	0.672997	0.672846	6.728463
63	Equatorial Guinea	0.372494	0.373063	0.372779	3.727788
64	Eritrea	0.489985	0.490421	0.490203	4.902028
65	Estonia	0.843896	0.844108	0.844002	8.440022
66	Ethiopia	0.495929	0.496358	0.496144	4.961438
67	Faeroe Islands	0.71571	0.715988	0.715849	7.158491
68	Falkland Islands Malvinas	0.339295	0.33991	0.339603	3.396027
69	Fiji	0.361065	0.361648	0.361357	3.613566
70	Finland	0.5	0.5	0.5	5
71	France	0.731409	0.731508	0.731459	7.314586
72	French Guiana	0.256926	0.257248	0.257087	2.570872
73	French Polynesia	0.502839	0.503265	0.503052	5.030518
74	Gabon	0.399632	0.40016	0.399896	3.998956
75	Gambia	0.173805	0.174703	0.174254	1.742539
76	Georgia	0.513447	0.513861	0.513654	5.136539
77	Germany	0.615415	0.615504	0.61546	6.154595
78	Ghana	0.660649	0.660868	0.660759	6.607585
79	Gibraltar	0.452123	0.452581	0.452352	4.523516
80	Greece	0.935115	0.935297	0.935206	9.352061
81	Greenland	0.321362	0.322	0.321681	3.216808



<b>index</b>	<b>country</b>	<b>left</b>	<b>right</b>	<b>middle</b>	<b>R0</b>
82	Grenada	0.355813	0.356402	0.356107	3.561074
83	Guadeloupe	0.711658	0.711943	0.7118	7.118002
84	Guatemala	0.571146	0.571516	0.571331	5.71331
85	Guinea	0.641914	0.642222	0.642068	6.420679
86	Guinea Bissau	0.419003	0.419361	0.419182	4.191824
87	Guyana	0.5	0.5	0.5	5
88	Haiti	0.373902	0.374293	0.374097	3.740973
89	Holy See	0.161943	0.163339	0.162641	1.626411
90	Honduras	0.659288	0.659506	0.659397	6.593971
91	Hungary	0.703281	0.703482	0.703382	7.033816
92	Iceland	0.944144	0.944322	0.944233	9.442327
93	India	0.5	0.5	0.5	5
94	Indonesia	0.883225	0.883363	0.883294	8.83294
95	Iran	0.5	0.5	0.5	5
96	Iraq	0.695555	0.69584	0.695697	6.956973
97	Ireland	1	1	1	10
98	Isle Of Man	0.78325	0.783495	0.783372	7.833724
99	Israel	0.774156	0.774395	0.774275	7.742752
100	Italy	1	1	1	10
101	Jamaica	0.521124	0.521531	0.521328	5.213276
102	Japan	0.289138	0.289232	0.289185	2.89185
103	Jordan	0.734565	0.734827	0.734696	7.346962
104	Kazakhstan	0.717486	0.717588	0.717537	7.17537
105	Kenya	0.698015	0.698296	0.698155	6.981554
106	Kuwait	0.442183	0.442396	0.44229	4.422895
107	Kyrgyzstan	0.662303	0.662439	0.662371	6.623713
108	Laos	0.298391	0.298856	0.298624	2.986237
109	Latvia	0.774831	0.775072	0.774952	7.749517
110	Lebanon	0.639236	0.639557	0.639397	6.393965
111	Lesotho	0.114481	0.1162	0.11534	1.153402
112	Liberia	0.389136	0.389665	0.389401	3.894008
113	Libya	0.511694	0.51211	0.511902	5.11902
114	Liechtenstein	0.54426	0.54465	0.544455	5.444549
115	Lithuania	0.401264	0.401812	0.401538	4.015379
116	Luxembourg	1	1	1	10
117	Macedonia	0.557347	0.557719	0.557533	5.575331
118	Madagascar	0.484963	0.48522	0.485092	4.850915
119	Malawi	0.296189	0.296573	0.296381	2.963811
120	Malaysia	0.209485	0.209708	0.209597	2.095967
121	Maldives	0.2797	0.280193	0.279947	2.799468
122	Mali	0.629204	0.629433	0.629318	6.293184

<b>index</b>	<b>country</b>	<b>left</b>	<b>right</b>	<b>middle</b>	<b>R0</b>
123	Malta	0.548145	0.548369	0.548257	5.482572
124	Marshall Islands	0.049898	0.056006	0.052952	0.52952
125	Martinique	0.509455	0.509751	0.509603	5.096032
126	Mauritania	0.221333	0.222128	0.22173	2.217303
127	Mauritius	0.658285	0.658466	0.658376	6.583756
128	Mayotte	0.727261	0.727531	0.727396	7.273961
129	Mexico	0.5	0.5	0.5	5
130	Micronesia	4.10E-11	4.60E-11	4.35E-11	4.35E-10
131	Moldova	0.836353	0.836569	0.836461	8.36461
132	Monaco	0.384405	0.384951	0.384678	3.846776
133	Mongolia	0.344302	0.344896	0.344599	3.44599
134	Montenegro	0.693475	0.693681	0.693578	6.935781
135	Montserrat	0.272757	0.273459	0.273108	2.73108
136	Morocco	0.769318	0.76956	0.769439	7.694391
137	Mozambique	0.425095	0.425588	0.425341	4.253414
138	Myanmar	0.466404	0.466723	0.466563	4.665634
139	Namibia	0.259132	0.259643	0.259388	2.593876
140	Nepal	0.5	0.5	0.5	5
141	Netherlands	1	1	1	10
142	New Caledonia	0.298897	0.299379	0.299138	2.99138
143	New Zealand	0.58401	0.584361	0.584186	5.841858
144	Nicaragua	0.256055	0.256783	0.256419	2.564195
145	Niger	0.926583	0.926765	0.926674	9.266741
146	Nigeria	0.401892	0.402407	0.40215	4.021495
147	Norway	1	1	1	10
148	Oman	0.371783	0.372173	0.371978	3.719781
149	Pakistan	0.790434	0.790598	0.790516	7.905162
150	Panama	1	1	1	10
151	Papua New Guinea	1.15E-05	1.67E-05	1.41E-05	0.000141
152	Paraguay	0.562177	0.562554	0.562365	5.623653
153	Peru	0.5	0.5	0.5	5
154	Philippines	0.5	0.5	0.5	5
155	Poland	1	1	1	10
156	Portugal	1	1	1	10
157	Qatar	0.95489	0.955069	0.95498	9.549798
158	Reunion	0.85251	0.852724	0.852617	8.526169
159	Romania	0.843222	0.843435	0.843328	8.433285
160	Russia	0.5	0.5	0.5	5
161	Rwanda	0.661518	0.661831	0.661675	6.616747
162	Saint Barthelemy	0.098543	0.100538	0.099541	0.995405
163	Saint Helena	0.0337	0.037912	0.035806	0.35806

<b>index</b>	<b>country</b>	<b>left</b>	<b>right</b>	<b>middle</b>	<b>R0</b>
164	Saint Kitts And Nevis	0.254206	0.254721	0.254463	2.544634
165	Saint Lucia	0.33452	0.335132	0.334826	3.348257
166	Saint Martin	0.113305	0.114415	0.11386	1.1386
167	Saint Pierre And Miquelon	4.47E-11	5.07E-11	4.77E-11	4.77E-10
168	Saint Vincent And The Grenadines	0.02245	0.027051	0.02475	0.247503
169	Samoa	0.026635	0.030482	0.028558	0.285581
170	San Marino	0.735086	0.735356	0.735221	7.352213
171	Sao Tome And Principe	0.012537	0.014626	0.013581	0.135815
172	Saudi Arabia	1	1	1	10
173	Senegal	0.629579	0.629904	0.629742	6.297416
174	Serbia	1	1	1	9.999998
175	Seychelles	0.214097	0.214668	0.214383	2.143825
176	Sierra Leone	0.565556	0.565922	0.565739	5.65739
177	Singapore	0.099436	0.099599	0.099518	0.995176
178	Sint Maarten	0.518107	0.518514	0.518311	5.183106
179	Slovakia	0.850243	0.850455	0.850349	8.503489
180	Slovenia	0.985832	0.986064	0.985948	9.859482
181	Solomon Islands	0.1771	0.17801	0.177555	1.775552
182	Somalia	0.274597	0.275292	0.274945	2.749447
183	South Africa	1	1	1	10
184	South Korea	0.897753	0.897798	0.897776	8.977756
185	South Sudan	0.366593	0.367171	0.366882	3.668817
186	Spain	0.449279	0.451625	0.450452	4.504519
187	Sri Lanka	0.5	0.5	0.5	5
188	State Of Palestine	0.33128	0.331512	0.331396	3.313961
189	Sudan	0.5	0.5	0.5	5
190	Suriname	0.302475	0.303123	0.302799	3.027989
191	Swaziland	0.296995	0.297661	0.297328	2.973283
192	Sweden	0.5	0.5	0.5	5
193	Switzerland	1	1	1	10
194	Syria	0.448907	0.449382	0.449144	4.491443
195	Taiwan	0.128989	0.129245	0.129117	1.291173
196	Tajikistan	0.750324	0.750389	0.750356	7.503563
197	Tanzania	0.432475	0.432972	0.432724	4.327238
198	Thailand	0.063397	0.063728	0.063562	0.635624
199	Timor Leste	0.003678	0.004774	0.004226	0.04226
200	Togo	0.43376	0.434252	0.434006	4.34006
201	Trinidad And Tobago	0.658831	0.659143	0.658987	6.58987
202	Tunisia	0.704734	0.705006	0.70487	7.048702
203	Turkey	1	1	1	10
204	Turks And Caicos Islands	0.299587	0.300255	0.299921	2.999208

index	country	left	right	middle	R0
205	UK	0.811092	0.811321	0.811207	8.112068
206	USA	0.533751	0.533861	0.533806	5.33806
207	Uganda	0.566348	0.566721	0.566534	5.665343
208	Ukraine	0.754558	0.754799	0.754679	7.546786
209	United Arab Emirates	0.285831	0.286145	0.285988	2.85988
210	Uruguay	0.659978	0.660136	0.660057	6.60057
211	Uzbekistan	0.667586	0.667736	0.667661	6.676609
212	Vanuatu	4.55E-11	5.16E-11	4.85E-11	4.85E-10
213	Venezuela	0.622237	0.622476	0.622357	6.223568
214	Viet Nam	0.053389	0.054157	0.053773	0.53773
215	Wallis And Futuna Islands	4.24E-11	4.78E-11	4.51E-11	4.51E-10
216	Western Sahara	0.043163	0.044571	0.043867	0.438674
217	Yemen	0.241754	0.242529	0.242142	2.421415
218	Zambia	0.412006	0.412373	0.412189	4.121894
219	Zimbabwe	0.34645	0.34704	0.346745	3.467451

#### 附录四：LSTM 感染人数预测结果

	Albania	Algeria	Andorra	Argentina	Armenia	Australia	Austria	Azerbaijan	Belarus	Belgium	...	Thailand
0	41.024541	990.861012	36.010510	14701.342523	229.562349	179.485167	488.894343	719.559192	836.684334	1846.341455	...	6782.200554
1	41.211638	820.226532	33.170011	14365.417902	230.514975	245.643192	495.295385	744.799438	833.127869	1849.949227	...	4099.701289
2	41.646713	570.988556	27.934991	14643.264121	231.967444	334.037469	502.903505	778.470603	829.302883	1857.918416	...	626.766480
3	42.587017	1645.956165	27.979016	11403.811335	233.959063	354.614815	510.090469	819.330461	825.675174	1874.687364	...	2881.443974
4	43.773978	1208.932423	30.536346	16340.642211	236.477650	339.061192	520.320164	862.518414	822.068310	1893.877389	...	47.918609
...	...	...	...	...	...	...	...	...	...	...	...	...
175	694.353500	578.855188	53.943572	9168.857600	669.960952	41.368168	2909.054368	815.773613	589.401819	4206.122935	...	76.838607
176	642.537204	570.571928	53.857142	9171.527041	668.556206	41.563612	2894.996093	813.279157	585.175690	4165.712099	...	76.894518
177	633.033693	565.574535	53.827196	9174.053873	668.727671	41.641800	2886.615595	807.972813	582.832174	4188.613425	...	76.936252
178	648.833905	562.988187	53.842058	9175.801464	669.593683	41.640389	2883.927778	800.820342	581.830796	4229.976283	...	76.959634
179	663.798674	561.593175	53.884723	9176.028538	670.061897	41.612489	2886.542883	793.117657	581.426529	4244.268156	...	76.959667

图 28 各国 LSTM 估计结果概览

