

考虑舆情的分时段 SIQR 模型的 新冠疫情分析与预测

鲍玉斌¹, 刘济霆¹, 李效宇²

(1. 东北大学 计算机科学与工程学院, 辽宁 沈阳 110169; 2. 东北大学 机器人科学与工程学院, 辽宁 沈阳 110169)

摘 要: 为了科学地研判本土新冠疫情复发时的发展趋势, 有效控制和防止疫情蔓延, 本文将疫情传播分为三个时段, 并综合考虑舆情与防控措施对各时段的干预效果, 建立了一种分时段具有非线性参数的新型冠状病毒肺炎传染模型. 在真实数据集上的仿真结果表明: 对感染者排查力度的快速增强会显著干预疫情扩散, 可使受感染人数增速减缓的拐点提前到来. 相比其他模型, 所提模型对大连 2020 年 12 月疫情的预测结果不仅均方误差小, 而且曲线的变化趋势更贴近现实情况. 此外, 本文对具有长自由传播期和强聚集性的疫情进行了仿真, 预测误差与同类模型相比较小. 这表明本模型具有一定的通用性与健壮性.

关 键 词: 新型冠状病毒肺炎; 传染病模型; 舆情因素; 仿真模拟; 预测

中图分类号: TP 391; O 29 **文献标志码:** A **文章编号:** 1005-3026(2022)08-1065-09

Time-Divided SIQR Model for COVID-19 Analysis and Prediction in Consideration of Public Opinion

BAO Yu-bin¹, LIU Ji-ting¹, LI Xiao-yu²

(1. School of Computer Science & Engineering, Northeastern University, Shenyang 110169, China; 2. School of Robot Science & Engineering, Northeastern University, Shenyang 110169, China. Corresponding author: BAO Yu-bin, E-mail: baoyubin@cse.neu.edu.cn)

Abstract: In order to scientifically identify and judge the trend of local COVID-19 recurrence, and to control and prevent the spread of its epidemic, a novel time-divided SIQR coronavirus pneumonia infection model with nonlinear parameters is established. It divides the epidemic spread into three periods, and comprehensively considers the intervention effect of public opinion and prevention measures on each period. The simulation results on several real datasets show that the rapid enhancement of the screening of infected and uncontrolled people will significantly interfere with the spread of the epidemic and make the inflection point of slowing down the number of infections come early. The prediction results of the epidemic situation at Dalian in December 2020 by the proposed model not only have smaller mean square error, but also the change trend of the predication curve is closer to the reality compared with other baseline models. Additionally, the epidemic situation with longer free transmission period and stronger aggregation is simulated by the proposed model. The prediction error is also smaller than that of the compared models on the real data. This shows that the model has certain universality and robustness.

Key words: COVID-19; epidemic model; public opinion factors; simulation; prediction

在全球疫情尚未结束背景下,我国仍然面临着因境外冷链病毒输入或境外感染人员输入等造成的小范围地区性的疫情复发.为有效防止疫情的扩散,各地政府主要采用了调查流行病学排

查密切接触者、通过新闻媒体对疫情信息进行快速广播以及大规模核酸检测,确定并隔离感染者的防控措施.许多学者针对疫情的预测展开了大量的研究工作,研究了疫情的传播特点和规律,预

收稿日期: 2021-07-29

基金项目: 国家自然科学基金资助项目(62072084).

作者简介: 鲍玉斌(1968-),男,吉林集安人,东北大学教授.

测结果也为决策者提供了参考信息.

目前国内外学者用多种手段仿真预测了新冠肺炎的传播过程,但预测模型多是基于 Kermack 等^[1]提出的 SIR(susceptible infected recovered) 仓室模型或 SEIR(susceptible exposed infected recovered) 仓室模型,并通过增加仓室或改变参数等方式进行优化.其中,周涛等^[2]用 SEIR 模型与 Wallinga 等^[3]创建的基本传染数经验估计公式得出基本病毒再生数 R_0 的取值区间为(2.8,3.3). Yang 等^[4]应用 SEIR 模型和 LSTM(long-short term memory) 模型预测出我国 2020 年的疫情高峰是在同年的 2 月下旬,并验证了 2020 年 1 月 23 日武汉封城对抑制疫情发展的重要意义.张琳^[5]分阶段拟合了国内疫情早期的确诊人数等数据,并对疫情传播作了预测. Kissler 等^[6]模拟了不同程度的社交隔离举措对降低疫情峰值和防止医疗系统崩溃的影响. Engbert 等^[7]在 SEIR 模型基础上,考虑了新冠肺炎的空间异质性,实现了准确的短期预测. Li 等^[8]考虑了有力的防控措施的影响,提出一种改进的 SEIQR(susceptible exposed infected quarantined recovered) 模型. Huang 等^[9]综合考虑环境因素和不同国家的控制措施,在 SIR 模型基础上建立新冠肺炎预测系统,并在西方国家的疫情预测上取得良好效果.

现有文献多用武汉疫情数据仿真国内疫情传播情形,未考虑在疫情防控常态化情况下病毒在城市内区域性传播,感染人数增长速度相对平稳等传播特点.此外,如今医疗资源充足,有条件展开大范围核酸检测以排查确诊感染者,并且疫情信息及时准确的披露使人们本能地减少外出,降低与潜在感染者的接触几率,从而一定程度上阻止了疫情的扩散.

为准确预测疫情传播并为管理人员提供有效的防控策略,本文考虑舆情因素与感染者的隔离措施对疫情传播的影响,并综合疫情传播各个时期的特点,建立了一种具有非线性参数的分时段的传染病模型——SIQR(susceptible infected quarantined recovered) 模型.它不仅对现阶段准确预测疫情传播有着重要意义,更为政府防疫部门科学有效地防控本土疫情提供决策参考.

1 模型建立与分析

1.1 模型建立

本文在经典的 SIR 模型基础上,将感染者加以隔离,增加了确诊并被隔离的感染者仓室 Q . 因

为不同的疫情传播时期,存在不同的主要干预措施与强度,故将感染率、隔离率和治愈率等参数分时期地描述,建立了非线性的 SIQR 模型,以得到更符合现实城市防疫特点的防疫模型.其中仓室 S 代表易感人群集合, I 代表具有感染性但没有被确诊的人群集合, Q 代表已被确诊并被隔离的感染者集合, R 代表已经治愈的感染者集合.模型仓室之间的状态转移图见图 1.

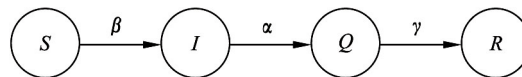


图 1 SIQR 模型状态转移图

Fig. 1 State transition diagram of the SIQR model

根据 SIQR 模型仓室的定义及状态转移图,可得 SIQR 模型的微分方程组如式(1)所示:

$$\left. \begin{aligned} \frac{dS(t)}{dt} &= -\frac{\beta \cdot S(t) \cdot I(t)}{N}, \\ \frac{dI(t)}{dt} &= \frac{\beta \cdot S(t) \cdot I(t)}{N} - \alpha \cdot I(t), \\ \frac{dQ(t)}{dt} &= \alpha \cdot I(t) - \gamma \cdot Q(t), \\ \frac{dR(t)}{dt} &= \gamma \cdot Q(t). \end{aligned} \right\} \quad (1)$$

其中,当易感人群 S 接触感染者人群 I 后,他们会以一定的感染率 β 成为感染者.隔离率 α 描述了防疫部门对感染者检测并隔离干预的强度.被隔离的感染者以一定的治愈率 γ 成为治愈者. N 代表发生疫情的城市人口总数.

1.2 感染率与舆情的关系

新冠肺炎的感染率 β 可以通过经验的基本病毒再生数 R_0 计算得到. R_0 的计算公式^[10] 见式(2):

$$R_0 = kbD. \quad (2)$$

其中: k 是单位时间内感染者的接触人数; b 是感染者能成功感染其接触的易感者的概率; D 是感染者传播病毒的平均时间.根据 R_0 , k 和 b 的定义,感染率 β 的计算公式如式(3)所示:

$$\beta = kb. \quad (3)$$

由于媒体对疫情新闻的报道,减少了人们不必要的外出活动,从而减小感染者的平均接触人数,即媒体对于感染率的影响主要反映在参数 k 上.

目前考虑媒体影响的模型多采用非线性的递减函数作为感染率函数.如 Xiao 等^[11]用负指数函数表达媒体影响,建立了 H1N1 流感传染病模型,并发现疫情早期媒体的影响最显著. Pawelek 等^[12]建立了考虑推特信息影响的流感模型,其媒

体影响函数见式(4):

$$\beta = \beta_0 \cdot e^{-\alpha T(t)}. \quad (4)$$

函数 $T(t)$ 描述了易感者、潜伏者和感染者以不同的系数接受信息,并且推特信息以一定速率变得过时的情形。

尽管该流感模型在 H1N1 流感仿真上表现良好,但不适合描述在疫情防控背景下短时间内的信息传播。此外,文献[11-12]主要从不同仓室受媒体影响的角度,提出了相应的模型。但本土疫情防控的新闻时效性强、受众广泛并且传媒方式丰富,首例病例报告当天同城民众基本都可得知疫情事件的发生。因此可以不考虑式(4)中函数 $T(t)$ 中不同人群以不同速率接受信息的差别,可对其进行简化。

综合上述考虑舆情影响的传染病模型的特点,本文采用负指数函数作为媒体影响函数。为客观且宏观地描述舆情,采用百度指数量化舆情,它既表示官方媒体对事件的重视程度,又表示民众对疫情事件的关注度。媒体舆情影响函数如式(5)所示:

$$\beta = \beta_0 \cdot e^{-nx}. \quad (5)$$

其中: x 代表归一化后的舆情数据变量; n 表示媒体对感染率影响的大小,是一个常量系数。

1.3 分时期的模型参数

根据复发的本土疫情实际传播情况,将疫情的传播分为三个时期:第一个时期为疫情传播自由期,零号病人具有感染性但未被确诊,造成了本土疫情传播;第二个时期为疫情传播初期,出现首例确诊病例,但尚未开展大规模核酸检测,并且感染率 β 受舆情的影响;第三个时期为疫情传播后期,对感染者的排查和隔离是干预疫情传播的最主要措施。

在疫情传播的自由期,防疫部门不知情本土疫情的出现,没有排查并收治感染者,所以设隔离率与治愈率均为 0,且感染率不受媒体的影响,是不变的。设这一时期的结束时刻是 t_1 ,则该时期的模型参数如式(6)所示:

$$\begin{cases} \alpha = \gamma = 0, \\ \beta = \beta_0, \end{cases} \quad 0 < t \leq t_1. \quad (6)$$

在疫情传播初期,防疫部门对于溯源传染病与排查感染者有一定的困难,故此时期隔离率 α 设为线性增大,以描述增长速度相对平稳的传播初期。感染率 β 设置为关于舆情变量的负指数函数。设疫情传播初期的结束时刻是 t_2 ,则此时期的模型参数如式(7)所示:

$$\begin{cases} \alpha = p \cdot (t - t_1) + q, \\ \beta = \beta_0 \cdot e^{-nx}, \end{cases} \quad t_1 < t \leq t_2. \quad (7)$$

其中, p, q 是根据疫情数据计算出来的参数。

由于复发的本土疫情传播周期短,故患者的治愈出院时间应该予以考虑,设为 Δt 。这段时间内少部分无症状感染者被解除观察隔离状态会使治愈病例人数增加,即首例病例出现后的 Δt 内治愈率很小,并假设之后治愈率随时间呈指数增长关系,则出现首例确诊病例后的治愈率见式(8):

$$\gamma = \begin{cases} \gamma_0, & t_1 < t \leq t_1 + \Delta t, \\ \gamma_0 \cdot e^{h \cdot (t - t_1 - \Delta t)}, & t_1 + \Delta t < t. \end{cases} \quad (8)$$

其中: t_1 为疫情传播自由期的结束时刻; h 为治愈率增高的快慢;治愈率的取值范围为 $\gamma_0 \leq \gamma \leq \gamma_{\max}$ 。

在疫情后期,民众对于疫情的关注度可能仍在升高,即媒体对于感染率仍存在一定的影响。当舆情指数达到峰值后,民众的防疫意识已然形成,故此后感染率不变($x = 1$)。设舆情指数峰值对应的时刻为 t_3 ,则疫情后期的感染率见式(9):

$$\beta = \begin{cases} \beta_0 \cdot e^{-nx}, & t_2 < t \leq t_3; \\ \beta_0 \cdot e^{-n}, & t > t_3. \end{cases} \quad (9)$$

如果疫情后期的隔离率不变,则疫情初期与疫情后期的隔离率大小相等,不符合实际防控过程中防控措施与力度随时间的推移不断增强的特点。因此,在疫情后期,为描述政府干预强度的快速增强,应用非线性递增的隔离率,并使疫情转折点滞后于疫情后期的开始时刻,见式(10):

$$\alpha(t) = \alpha_0 \cdot e^{m(t - t_2)}, \quad \alpha_0 \leq \alpha(t) \leq \alpha_{\max}. \quad (10)$$

其中,参数 m 描述了隔离干预程度增强的快慢,取值范围通常在(0.15, 0.25)。

2 隔离率参数求解与讨论

2.1 疫情传播初期的隔离率参数求解

在疫情传播的自由期,没有对感染者排查的干预措施,即没有仓室 I 到仓室 Q 的转移过程,此时期是经典的 SI(susceptible infected)模型,感染者的微分方程见式(11):

$$\frac{dI(t)}{dt} = \frac{\beta \cdot S(t) \cdot I(t)}{N}. \quad (11)$$

易感者人数 $S(t)$ 与城市总人数 N 的差非常小,可以认为它们的比值约为 1,对时间 t 积分可得感染者人数关于 t 的函数见式(12):

$$I(t) = I_0 \cdot e^{\beta \cdot (t - t_1)}, \quad 0 < t \leq t_1. \quad (12)$$

其中: I_0 是初始感染者人数; t_1 是疫情传播自由期的结束时刻,则在 t_1 时刻的感染者人数如式(13)所示:

$$I(t_1) = I_0 \cdot e^{\beta \cdot (t_1 - t_1)}. \quad (13)$$

在疫情传播初期,同样认为易感者人数与总人数的比值约为 1,则感染者仓室的微分方程简化成式(14):

$$\frac{dI(t)}{dt} = \frac{\beta \cdot S(t) \cdot I(t)}{N} - \alpha \cdot I(t) \approx (\beta - \alpha) \cdot I(t) \quad (14)$$

设 t_2 是此时期的结束时刻,则式(14)的解为

$$I(t) = I(t_1) \cdot e^{(\beta - \alpha) \cdot (t - t_1)}, \quad t_1 < t \leq t_2. \quad (15)$$

模型中仓室 Q 定义为被确诊的感染者集合,所以官方统计的确诊人数数据包括被确诊的感染者和已被治愈的感染者,即累计确诊人数 $N(t)$ 的表达式如式(16)所示:

$$N(t) = Q(t) + R(t). \quad (16)$$

对式(16)左右两端同时求导,得到疫情初期累计确诊人数的微分方程,如式(17)所示:

$$\frac{dN(t)}{dt} = \frac{dQ(t)}{dt} + \frac{dR(t)}{dt} = \alpha \cdot I(t). \quad (17)$$

将式(15)代入式(17),得到式(18):

$$\frac{dN(t)}{dt} = \alpha \cdot I(t_1) \cdot e^{(\beta - \alpha) \cdot (t - t_1)}. \quad (18)$$

对式(18)左右两端同时积分,得疫情传播初期的确诊人数函数 $N(t)$,如式(19)所示:

$$N(t) = \frac{\alpha}{\beta - \alpha} \cdot I(t_1) \cdot (e^{(\beta - \alpha) \cdot (t - t_1)} - 1). \quad (19)$$

其中,隔离率 α 是式(7)中定义的关于时间 t ($t_1 < t \leq t_2$) 的线性函数.

通过式(19)对真实确诊人数的数据进行拟合,得到参数 p 和 q ,将其代入式(7),即可求得疫情传播初期的隔离率.

2.2 疫情后期隔离率参数讨论

对参数 m 和 α_{\max} 进行对比实验,根据大连市疫情防控工作公布的传染病流行病学调查结果,可以认为疫情传播自由期为 5 d,并根据大连市采取排查感染者强度的不同,设疫情传播初期为 5 d.大连市疫情自 12 月 15 日至次年 1 月 30 日共 47 d 的确诊人数数据在辽宁省卫健委官方网站可得^①.从百度指数官网^②得到大连市疫情初期的舆情数据.

设归一化后舆情数据变量 $x=0$ 时有平均接触人数 $k_1=10$ ^[13],当 $x=1$ 时有平均接触人数 $k_2=8$.根据式(3)和式(5)易求得参数 n :

$$n = \ln\left(\frac{k_1}{k_2}\right). \quad (20)$$

根据文献[2],设 R_0 为 3.3,并设置感染者传播病毒的平均时间 $t=7$ d,根据式(2),式(3)可以求得 β_0 .综上,对比实验参数设置如表 1 所示.

表 1 除 m 和 α_{\max} 外的模型参数
Table 1 The model parameters except m and α_{\max}

参数	β_0	n	p	q
数值	0.471	0.223	0.04	0.05

为讨论疫情传播后期隔离率 $\alpha(t)$ 计算(见式(10))中参数 m 和 α_{\max} 的效用,设参数 α_{\max} 为 0.6,仿真不同参数 m 时累计确诊人数随时间的变化情况,如图 2a 所示.同理,设参数 m 为 0.2,仿真不同参数 α_{\max} 时累计确诊人数随时间的变化情况,见图 2b.

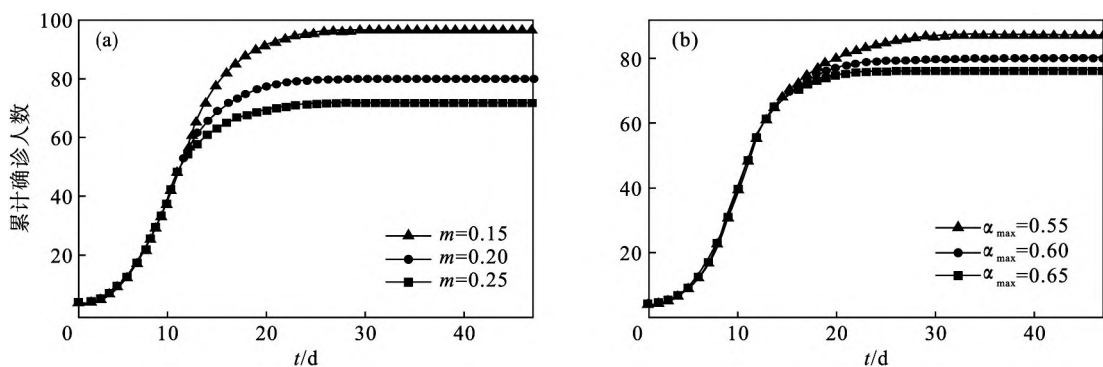


图 2 参数 m 与 α_{\max} 分别为控制变量的对比实验
Fig. 2 The comparative experiments when parameters m and α_{\max} as controlled variables respectively

(a) —参数 m ; (b) —参数 α_{\max} .

由图 2 可以看出,参数 m 的变化使得增长曲线的变化较大.相比之下,参数 α_{\max} 的变化却没有

使增长曲线表现出较大的变化.从最后的确诊总人数看, m 值的变化使其有明显的差异,而 α_{\max} 的

① <http://wjw.shenyang.gov.cn/>

② <https://index.baidu.com/>

不同使曲线的平缓趋势有所不同. 因此参数 m 主要决定曲线的增长趋势, 而参数 α_{\max} 主要决定着曲线最终的平稳趋势.

疫情转折点是衡量疫情控制效果的关键, 需要定量地计算. 本模型中感染者人数 I 的微分方程为

$$\frac{dI(t)}{dt} = \frac{\beta \cdot S(t) \cdot I(t)}{N} - \alpha \cdot I(t) \quad (21)$$

疫情的转折点为感染者人数开始减少的时刻, 即式 (21) 小于 0 的时刻, 故此时的隔离率为

$$\alpha > \frac{\beta \cdot S(t)}{N} \quad (22)$$

为定量地说明参数 m 和 α_{\max} 对疫情防控的干预程度, 根据式 (22), 计算得到图 2a 对应实验中参数 m 不同取值条件下的疫情转折点, 如表 2 所示. 当 $m = 0.2$ 时, α_{\max} 不同取值的疫情转折点如表 3 所示.

表 2 参数 m 取值对疫情转折点的影响
Table 2 The impact of m value on the turning point of the epidemic situation

m	转折点 /d	确诊总人数
0.15	10	97
0.20	9	80
0.25	8	72

表 3 参数 α_{\max} 取值对疫情转折点的影响
Table 3 The impact of α_{\max} value on the turning point of the epidemic situation

α_{\max}	转折点 /d	确诊总人数
0.55	9	88
0.60	9	80
0.65	9	76

表 3 中 α_{\max} 取值不同时的疫情转折点相同, 说明一定范围内变化的 α_{\max} 对于疫情干预影响较弱, 仅使确诊总人数有一定的差异. 但是表 2 显示参数 m 的变化将疫情转折点时间提前, 使得疫情提前控制, 最终感染总人数也会减少. 因此, 防疫部门对疫情的反应速度一定程度上决定了最终的感染总人数.

3 实验与结果分析

3.1 预测模型评估与对比

3.1.1 SIQR 模型预测累计确诊人数

从预测与决策的角度看, 舆情指数的峰值选择为首例确诊病例确诊后 8 d 内的高峰值, 而时段的划分是需要决策者根据疫情发生的实际情况

作出判断, 如发生地是否为人员密集流动区, 传染链是否已存在较长时间等. 根据不同的疫情转折时间点, 用疫情传播初期的数据可以预测最终累计确诊人数的范围.

实验数据为 2.2 节提及的大连市疫情数据与舆情数据. 根据传染病流行病学调查结果及大连市疫情防控工作报告, 设疫情传播自由期和传播初期均为 5 d. 基本再生数 R_0 沿用 2.2 节中的取值, 分时期的感染率、隔离率用式 (6), (7), (9), (10) 计算. 累计确诊人数是隔离者和治愈者数量之和, 故不需考虑治愈率参数.

疫情传播初期的隔离率公式 (见式 (7)) 中的参数 p 和 q 由式 (19) 对 12 月 15 号至 12 月 17 日的确诊人数进行有约束的拟合得到, 拟合工具为 Python 及 Imfit 包. 由 2.2 节的讨论可知, α_{\max} 不影响转折时间点, 因此主要讨论不同疫情转折点下的参数 m 取值范围和最终感染人数等, 本模型的部分实验参数值见表 4.

表 4 模型主要参数
Table 4 Model parameters of simulation experiment

参数	β_0	n	p	q	α_{\max}
数值	0.471	0.223	0.047	0	0.60

为描述疫情得到控制后的稳定状态, 定义疫情平稳时间为连续 3 d 确诊人数的增长值小于 1, 即

$$N(t_i) - N(t_{i-1}) < 1 \quad (23)$$

根据实际情况设置疫情转折点为出现首例确诊病例后的第 8 天至第 10 天. 据此计算 m 的取值范围、总感染人数以及平稳期, 结果见表 5.

表 5 不同转折点条件下的预测结果
Table 5 The predication results at different turning points

疫情转折点	第 8 天	第 9 天	第 10 天
m 取值范围	0.24 ~ 0.34	0.18 ~ 0.23	0.14 ~ 0.17
最终感染人数	66 ~ 76	77 ~ 90	93 ~ 107
平稳期 /d	20 ~ 22	22 ~ 23	24 ~ 25

由表 5 可知, 参数 m 的取值随着疫情转折点的推移而减小, 并且最终的感染人数与平稳期的数值均有明显的分段. 如果 m 取值为区间的中间值 0.20, 当疫情在第 9 天达到转折点, 且最终确诊人数预计约为 85 时, 预测曲线如图 3 所示.

3.1.2 不同预测模型对比

为了说明本模型更适用于预测“后疫情”时代本土疫情变化趋势, 将其与 Logistic 模型, Logistic 与 SEIR 结合模型 (Logistic - SEIR)^[14] 相

对比. 文献 [14] 对 Logistic - SEIR 模型的参数进行了讨论, SEIR 模型参数 $S_0 = 1.35K_0$, E_0 用 Logistic 模型中的参数 P_0 初始化, I_0 设为第 1 天的感染人数, 传染率初值设为 0.25.

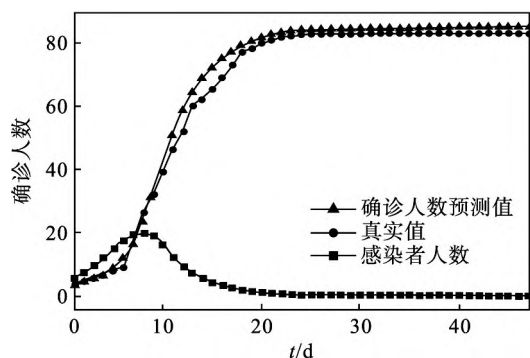


图 3 大连市疫情累计确诊人数的预测结果

Fig. 3 Forecast results of cumulative confirmed cases in Dalian

因为 Logistic 模型只能预测累计确诊人数, 而且本文模型的治愈率参数较难从预测角度确定, 故上述三种模型的预测对象为累计的确诊人数. 上述两种对比模型的数据集依据已知确诊人数的天数来划分, 分别为 10, 13, 16 d, 预测结果见表 6. 本模型的已知天数为 8d, 评价指标采用均方根误差 RMSE, 即

$$RMSE = \sqrt{\frac{1}{\tau} \sum_{i=1}^{\tau} (p_i - t_i)^2}. \quad (24)$$

其中: p_i 和 t_i 分别是第 i 个时刻的仿真值和真实值; τ 表示疫情传播全时期的天数.

表 6 不同模型的预测性能

Table 6 The predictive performance of the different models

预测模型	已知确诊人数的天数	最终确诊人数预测值	RMSE
本模型	8 ($m = 0.16$)	98	12.71
	8 ($m = 0.20$)	85	2.94
	8 ($m = 0.25$)	76	5.97
Logistic	10	102	15.78
	13	76	5.27
	16	73	7.68
Logistic - SEIR	10	112	23.04
	13	87	19.94
	16	84	19.85

由表 6 可知, 本模型与 Logistic 模型的预测范围接近, 而 Logistic - SEIR 的预测范围比实际值偏高. 本模型需要的已知天数较少, 但需要决策者对疫情转折点作出判断, 而 Logistic 模型在已知天数较少时 (如 8 d), 预测的最终确诊人数值

会远超实际值. Logistic - SEIR 模型也是如此.

本模型 ($m = 0.20$)、Logistic 模型和 Logistic - SEIR 模型 (已知天数都为 13 d) 的累计确诊人数见图 4.

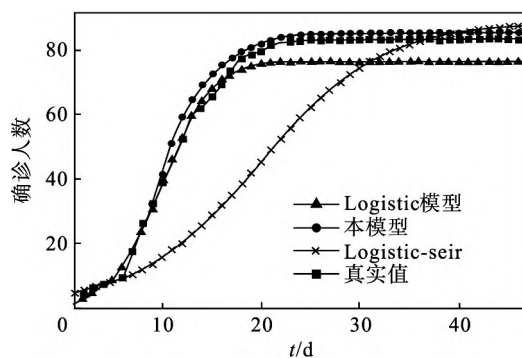


图 4 不同模型的预测性能

Fig. 4 Predictive performance of different models

本模型与 Logistic 模型预测的疫情变化趋势与实际一致, 而 Logistic - SEIR 模型的预测结果与实际不符.

综上, 本模型可以较好地预测复发的本土疫情累计确诊人数与疫情发展的变化趋势.

3.2 考虑舆情对 SIQR 模型的改进效果

为了说明考虑舆情的 SIQR 模型的效果, 将其与感染率 β_0 不变的 SIQR 模型分别拟合大连市 2020 年 12 月真实的疫情确诊人数曲线, 进行对比实验. 因为本文实验 $t=0$ 是首例病例的确诊时间, 考虑患者治愈出院时间对模型的影响, 设首例病例出现后 15 d 内, 治愈病例数量很少, 故此时间段内的治愈率很小. 文献 [14] 对治愈率研究的结论是不断增大治愈率会使模型表现更好, 则公式 (8) 的参数取值为 $\gamma_0 = 0.01$, $h = 1/5$ 和 $\gamma_{\max} = 0.5$, $m = 0.21$, 其余参数设置见表 4.

对比实验仅感染率定义不同, 其余参数均相同, 得到疫情变化的仿真曲线如图 5a 和图 5b 所示. 非线性感染率模型的仓室 Q 与 R 的仿真曲线如图 5c 和图 5d 所示, $Q(t)$ 与 $R(t)$ 的值分别为现有确诊人数和累计治愈病例数.

从图 5a 中可以看出考虑舆情的 SIQR 模型仿真结果与真实的确诊人数变化趋势相符, 而图 5b 中未考虑舆情的曲线趋势远超过了实际的变化过程, 误差较大. 尽管增大隔离率参数可以使得最终确诊人数与实际相近, 但确诊人数的增长速度将非常快, 与实际不符. 图 5c 和图 5d 中的 $Q(t)$ 和 $R(t)$ 仿真曲线均与真实的变化趋势一致, 其与真实值的均方根误差分别为 3.48 和 3.63, 进一步说明了本模型的仿真结果较好. 应用本模型在考虑舆情与未考虑两种情形下进行仿真, 结果见表 7.

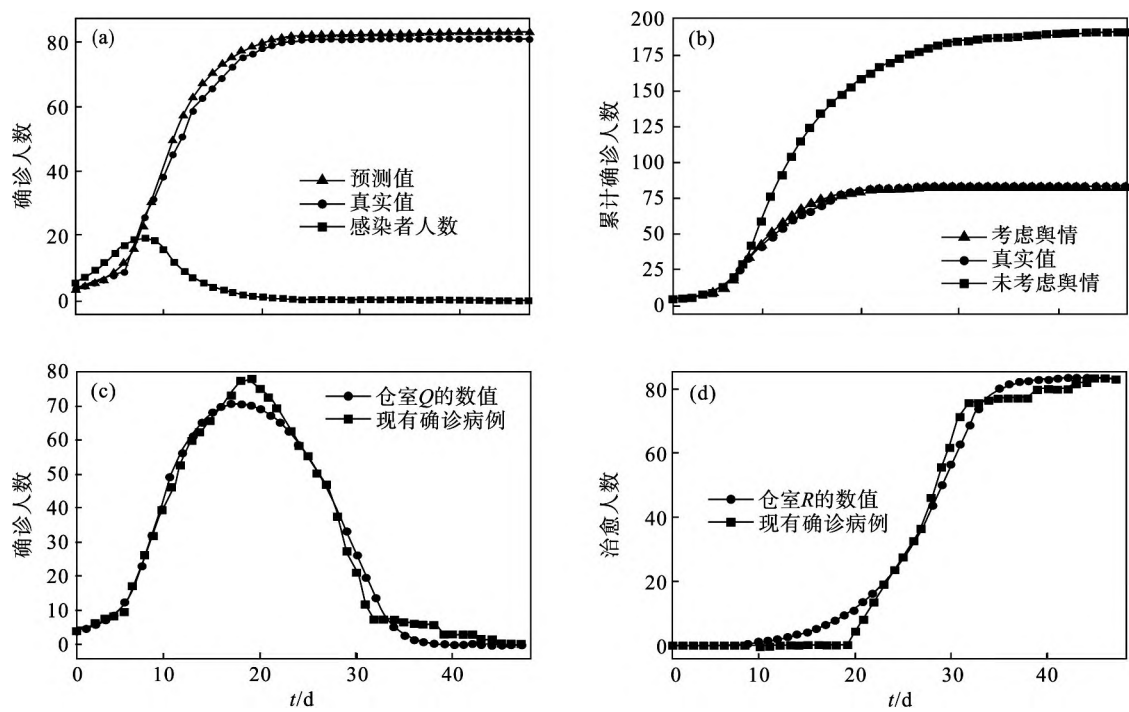


图 5 大连市疫情仿真结果
Fig. 5 Simulation results of epidemic in Dalian
(a) 一非线性感染率的仿真结果; (b) 一考虑舆情与否的结果对比; (c) 一考虑舆情的 SIQR 模型的 $Q(t)$ 仿真曲线;
(d) 一考虑舆情的 SIQR 模型的 $R(t)$ 仿真曲线.

表 7 考虑舆情与未考虑舆情的模型仿真结果 Table 7 Simulation results of models considering and not considering public opinion		
变量	考虑舆情的仿真	未考虑舆情的仿真
RMSE	2.13	81.60
转折点/d	9	10

表 7 中考虑舆情的仿真误差仅为 2.13,说明仿真结果与真实值相近. 疫情转折点在疫情后期第 9 天到来,从确诊人数的数据上看,也比较符合真实的情况. 而未考虑舆情的仿真误差值较大.

因此,本文所提出的考虑舆情的 SIQR 模型不仅预测准确度较高,而且能够更好地拟合疫情发展的变化趋势,改进效果明显.

3.3 对传播自由期较长的疫情的仿真模拟

通常情况下,复发的本土疫情具有确诊人数增长较为平稳的特点. 但也不乏例外,如 2020 年 6 月北京新发地疫情以及 2021 年初河北省疫情. 这类疫情往往具有传播自由期较长,发生过聚集性感染以及疫情防控响应速度快等特点. 因此,为说明本文模型的健壮性与实用性,现对北京新发地疫情进行全时期的仿真模拟.

从北京市卫健委官网^③,收集了北京市 2020

年 6 月 11 日至 8 月 6 日共 57 d 的确诊病例和治愈病例数据. 在百度指数官网收集了关于“北京疫情”关键字的舆情数据,相比于大连 12 月疫情时的舆情数据,北京的舆情指数远高于大连的,说明北京疫情扩散受媒体影响更大. 所以设 $x=0$ 时的平均接触人数 $k_1=10$,当 $x=1$ 时平均接触人数 $k_2=7$. 根据式(20)可以求得参数 n .

根据北京市卫健委公布的流行病学调查结果,认为 5 月 31 日为疫情传播自由期的开始时间,共 11 d. 由于防控部门响应迅速,并且日均检测量高,故疫情初期为 2 d. 初期第 1 天出现确诊病例,只需对第 2 天的隔离率 α_0 作出设置(α_0 是疫情初期最后一天隔离率的值),故无需再拟合求隔离率. 根据实际防疫强度, α_0 可以在区间 $[0.1, 0.2]$ 之间取值,使疫情确诊人数的增长曲线较为贴合实际. 由于北京新发地属于人员聚集区,故设 R_0 为 3.5,病毒传播的平均时间 D 不变. 参考 3.2 节治愈率参数的设置,取 $\Delta t=20$, $\gamma_0=0.01$, $h=1/7$ 和 $\gamma_{\max}=0.5$. 其余参数见表 8.

根据表 8 中的模型参数,用本文模型对北京市 2020 年 6 月疫情进行全时期仿真,得到累计确诊人数的拟合曲线与潜在感染者人数的估计曲线见图 6a,仿真中仓室 Q 与 R 的变化曲线见图 6c

③ <http://wjw.beijing.gov.cn>

和图 6d. 为与本文模型的拟合效果作对比,用 Logistic 模型同样进行全时期的仿真模拟,结果如

表 8 仿真实验的模型参数
Table 8 Model parameters of simulation experiment

参数	β_0	α_0	α_{\max}	m	γ_0	h	γ_{\max}
数值	0.50	0.20	0.50	0.18	0.01	1/7	0.50

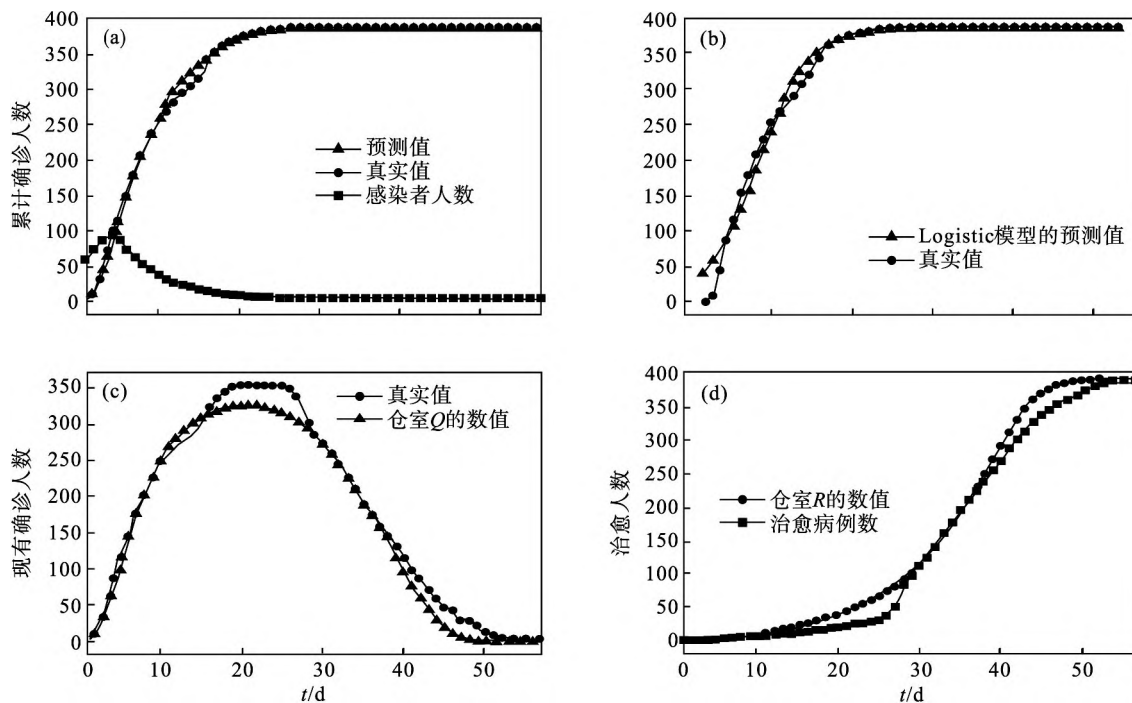


图 6 北京市疫情仿真结果

Fig. 6 Simulation results of epidemic in Beijing

(a) 一本文模型的仿真结果; (b) 一Logistic 模型的仿真结果; (c) 一本文模型的 $Q(t)$ 曲线; (d) 一本文模型的 $R(t)$ 曲线.

用式(22) 计算得到本模型疫情仿真结果的转折点,则本模型与 Logistic 模型对北京市疫情进行仿真,结果见表 9.

表 9 北京市疫情仿真结果
Table 9 Simulation results of Beijing epidemic

模型	RMSE	转折点/d	平稳时刻/d
本模型	8.00	5	32
Logistic 模型	11.64	—	27

从图 6a 可以看出该模型得到预测曲线的变化趋势与实际确诊人数的变化趋势一致,预测值与真实值相近. 根据仿真结果,在首例病例确诊后第 5 天即为疫情的转折点,可知北京市对这次疫情的防控强度大,反应速度快,但是由于病毒已传播一段时间,并且源头为人员密集地区,所以造成了较多人员的感染. 模型仿真值与真实值的均方根误差为 8.00,且比 Logistic 模型的误差小,并且平稳时刻的估计值更符合实际. 相较于确诊

人数的量级,预测值与真实值误差不大. 但本文模型的 $Q(t)$ 和 $R(t)$ 曲线与真实的误差分别为 18.59 和 18.04,误差较大的原因在于现有确诊人数的峰值高且处于峰值时有段时间数值变化很小.

综上,本模型对于疫情传播自由期长,存在聚集性感染情况的本土复发疫情也有较好的拟合效果,因此,模型的适应性与健壮性较好.

4 结 语

本文针对本土疫情防控的特点,将舆情因素纳入模型,提出了分时段的 COVID-19 疫情预测 SIQR 模型,并在实际数据上进行了对比测试,验证了本文模型具有较好的预测性能和鲁棒性. 本文的模型可以通过选择不同的疫情转折点,对

(下转第 1088 页)

- mathematical models and their Monte Carlo estimates [J]. *Mathematics and Computers in Simulation*, 2001, 55 (1/2/3) : 271 – 280.
- [12] Sudret B. Global sensitivity analysis using polynomial chaos expansions [J]. *Reliability Engineering & System Safety*, 2008, 93(7) : 964 – 979.
- [13] Quicken S, Donders W P, van Disseldorp E M J, et al. Application of an adaptive polynomial chaos expansion on computationally expensive three-dimensional cardiovascular models for uncertainty quantification and sensitivity analysis [J]. *Journal of Biomechanical Engineering*, 2016, 138(12) : 121010.
- [14] Bratley P, Fox B L. Implementing Sobols quasirandom sequence generator(algorithm 659) [J]. *ACM Transactions on Mathematical Software*, 2003, 29(1) : 49 – 57.
- [15] Willemet M, Chowienzyk P, Alastruey J. A database of virtual healthy subjects to assess the accuracy of foot-to-foot pulse wave velocities for estimation of aortic stiffness [J]. *American Journal of Physiology-Heart and Circulatory Physiology*, 2015, 309(4) : H663 – H675.

(上接第 1072 页)

全时期疫情变化趋势以及确诊总人数作出相应的预测,可为决策者提供参考.根据实验结果分析发现,如果由转折点计算得到的 m 值在 (0.15, 0.25) 之间,那么预测的结果较好.全时期疫情数据的拟合结果表明,用舆情数据修正得到的感染率能较好地适应各时期疫情扩散的特点,使全时期疫情传播的仿真结果更佳.

参考文献:

- [1] Kermack W O, McKendrick A G. A contribution to the mathematical theory of epidemics [J]. *Proceedings of Royal Society of London, Series A*, 1927, 115(772) : 700 – 721.
- [2] 周涛,刘权辉,杨紫陌,等.新型冠状病毒肺炎基本再生数的初步预测 [J]. *中国循证医学杂志*, 2020, 20(3) : 359 – 364.
(Zhou Tao, Liu Quan-hui, Yang Zi-mo, et al. Preliminary prediction of the basic reproduction number of the novel coronavirus 2019-nCoV [J]. *Journal of Evidence-Based Medicine*, 2020, 20(3) : 359 – 364.)
- [3] Wallinga J, Lipsitch M. How generation intervals shape the relationship between growth rates and reproductive numbers [J]. *Proceedings Biological Sciences*, 2007, 274 (1609) : 599 – 604.
- [4] Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions [J]. *Journal of Thoracic Disease*, 2020, 12(3) : 165 – 174.
- [5] 张琳.新冠肺炎疫情传播的一般增长模型拟合与预测 [J]. *电子科技大学学报*, 2020, 49(3) : 345 – 348.
(Zhang Lin. Fitness of the generalized growth to the COVID-19 data [J]. *Journal of University of Electronic Science and Technology of China*, 2020, 49(3) : 345 – 348.)
- [6] Kissler S M, Tedijanto C, Goldstein E, et al. Projecting the transmission dynamics of SARS-CoV-2 through the post pandemic period [J]. *Science*, 2020, 368(6493) : 860 – 868.
- [7] Engbert R, Rabe M M, Kliegl R, et al. Sequential data assimilation of the stochastic SEIR epidemic model for regional COVID-19 dynamics [J]. *Bulletin of Mathematical Biology*, 2021, 83(1) : 1 – 16.
- [8] Li B, Liu Z, Zhao C, et al. A susceptible-exposed-infected-quarantined-recovered (SEIQR) model for predicting the trajectory of the COVID-19 epidemic [C] // Chinese Automation Congress(CAC). Shanghai: Chinese Association of Automation, 2020: 58 – 63.
- [9] Huang J P, Zhang L, Liu X Y, et al. Global prediction system for COVID-19 pandemic [J]. *Science Bulletin*, 2020, 65 (22) : 1884 – 1887.
- [10] Lipsitch M, Marc T, Cohen B, et al. Transmission dynamics and control of severe acute respiratory syndrome [J]. *Science*, 2003, 300(5627) : 1966 – 1970.
- [11] Xiao Y, Tang S, Wu J. Media impact switching surface during an infectious disease outbreak [J]. *Scientific Reports*, 2015, 5(1) : 1 – 9.
- [12] Pawelek K A, Oeldorf-Hirsch A, Rong L. Modeling the impact of Twitter on influenza epidemics [J]. *Mathematical Biosciences & Engineering*, 2014, 11(6) : 1337 – 1356.
- [13] 付强,姚羽.基于多隔离策略的新冠肺炎疫情影响建模及分析 [J]. *东北大学学报(自然科学版)*, 2020, 41(9) : 1239 – 1243.
(Fu Qiang, Yao Yu. Modeling and analysis of COVID-19 based on multi-isolation strategy [J]. *Journal of Northeastern University(Natural Science)*, 2020, 41(9) : 1239 – 1243.)
- [14] 冯苗胜,王连生,林文水. Logistic 与 SEIR 结合模型预测新型冠状病毒肺炎传播规律 [J]. *厦门大学学报(自然科学版)*, 2020, 59(6) : 1041 – 1046.
(Feng Miao-sheng, Wang Lian-sheng, Lin Wen-shui. Prediction of COVID-19 transmission law by combining logistic and SEIR model [J]. *Journal of Xiamen University (Natural Science)*, 2020, 59(6) : 1041 – 1046.)