一类 SEIR-A 与 TCN 混合传染病模型的研究

李 季,乔 敏,邹黎敏
重庆工商大学 数学与统计学院,重庆 400067

摘 要:目的 由于传染病的不断复杂化,一类传播过程中出现大量隐性感染人群的疾病逐渐流行,提出一种基于 SEIR 模型改进后的 SEIR-A 模型来更加准确地刻画该类疾病的传播机制。**方法** 在动力学模型方面,主要有以下两点改进:一是假设潜伏期人群和显性感染人群具有一致的传染因子;二是引入具有不同传染性的隐性感染人群 A,且增添隐性感染者向显性感染者单向转换的常值因子 ω,构建一类具有特殊隐性感染人群的 SEIR-A 模型。此外,将改进后的 SEIR-A 模型与时间卷积网络 TCN 模型线性结合,得到一种动力学模型和深度学习模型相互融合的混合模型。**结果** 通过真实数据的拟合,结果表明:SEIR-A 模型可以模拟传染病的总体趋势,且能够对该疾病中现存隐性感染人群和累计恢复人群做出准确拟合,决定系数 R^2 分别达到 0.987 0 和 0.989 9,证明该模型合理;SEIR-A 与 TCN 的混合模型可以实现对复杂现存显性感染人群的拟合,相较于单一的 SEIR-A 模型、TCN 和 LSTM 模型,该混合模型的决定系数 R^2 达到了 0.961 1,取得了 5 种对比模型中最优的拟合精度。**结论** 传统动力学和深度学习的结合,可以在体现疾病传播机理的同时有效解决传统模型拟合精度不高的问题,对传染病研究具有现实意义。

关键词:SEIR-A 模型;TCN 模型;粒子群优化算法;混合模型
中图分类号:O211.9 **文献标识码:**A doi:10.16055/j.issn.1672-058X.2023.0006.011

Study on a Mixed Epidemic Model of SEIR-A and TCN

LI Ji, QIAO Min, ZOU Limin
School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China

Abstract: Objective A modified SEIR-A model based on the SEIR model was proposed to describe the transmission mechanism of infectious diseases more accurately, in view of the increasing complexity of infectious diseases and the gradual prevalence of a kind of diseases with a large number of latent infections in the process of transmission. **Methods** In terms of the kinetic model, there were two main improvements. One was to assume that the incubation period population and the dominant infected population had the same infectious factors; the other was to introduce a recessive infection population A with different infectiousness, and add a constant factor ω for one-way conversion from recessive infection to dominant infection, so as to construct an SEIR-A model of a special recessive infection population. In addition, the improved SEIR-A model was linearly combined with the TCN model of time convolution network, and a hybrid model with the fusion of dynamic model and deep learning model was obtained. **Results** Through the fitting of real data, the results showed that the SEIR-A model could simulate the general trend of infectious diseases, and could accurately fit the existing recessive infected population and the accumulated recovered population in the disease, and the determination coefficient R^2 reached 0.987 0 and 0.989 9, respectively, proving that the model was reasonable. The hybrid model of SEIR-A and TCN can fit the complex existing dominant infected population. Compared with the single

收稿日期:2022-09-12 **修回日期:**2022-10-19 **文章编号:**1672-058X(2023)06-0083-10
基金项目:重庆市教育委员会项目(KJQN202000816).
作者简介:李季(1987—),男,湖北宜昌人,讲师,从事图像分割研究.
通讯作者:乔敏(1999—),女,甘肃天水人,硕士研究生,从事大数据分析研究. Email:1519146945@qq.com.
引用格式:李季,乔敏,邹黎敏. 一类 SEIR-A 与 TCN 混合传染病模型的研究[J]. 重庆工商大学学报(自然科学版),2023,40(6):83—92.
LI Ji, QIAO Min, ZOU Limin. Study on a mixed epidemic model of SEIR-A and TCN[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(6): 83—92.

SEIR-A model, TCN model and LSTM model, the determination coefficient R^2 of the hybrid model reached 0.9611, obtaining the best fitting accuracy among the five comparative models. **Conclusion** The combination of traditional dynamics and deep learning can effectively solve the problem of low fitting accuracy of traditional models while embodying the mechanism of disease transmission, which has practical significance for the study of infectious diseases.

Keywords: SEIR-A model; TCN model; particle swarm optimization algorithm; hybrid model

1 引 言

由于传染病在全球不断出现,在研究传染病机理的过程中,发现一类包含潜伏期的 SEIR 模型出现了局限性,如病毒在潜伏期间也可能存在传染性,且患者在经历潜伏期之后可能被感染但未出现相关症状的隐性情况。针对这一类现象,需要建立更为复杂的动力学模型。李峰[1]、TANG Y L[2]、马千里等[3]都针对隐性感染者构建了不同的动力学模型,来分析疾病的传播机制。虽然这些含有隐性感染人群的模型符合传染病的发展规律,但较为刻板,都不能准确体现隐性感染者与显性感染者之间的转换。

此外,将传统动力学与机器学习融合进行建模的方法也被广泛应用。冯苗胜等[4]提出一种将统计学与动力学融合的 Logistic_SEIR 模型,解决了单一模型需多次调参的局限; Malavika 等[5]使用 Logistic 增长曲线模型、SIR 模型对印度和部分高发州新患者进行了预测;张晴[6]采用得 SIR 模型、Logistic 模型相结合的方式,对传染病的传播情况进行研究和预测;谢小良等[7]提出一种 SEIR 模型和 ARIMA 模型搭建的 SEIR-ARIMA 混合模型。上述学者都采用混合模型的形式来研究传染病,这些方法在大数据时代下都可以取得较高的预测精度,但对传播机理的解释不够完善。

基于上述,目前还没有综合考虑疾病潜伏期传染性和隐性感染人群向显性感染人群单向流动的动力学模型,因此,本文在基础 SEIR 模型中假设潜伏期具有传染性,同时引入隐性感染人群,增添隐性与显性感染人群之间的转换因子作为影响参数,构建一种含隐性感染人群的 SEIR-A 模型。此外,通过构建 SEIR-A 与 TCN 的混合模型,有效减少不可控因素对疾病传播的影响,弥补了原有数学模型的不足。

2 研究方法

首先,文章基于经典的 SEIR 模型,将隐性感染人群加入传统的动力学模型中,同时考虑潜伏期的传染性,得到改进的 SEIR-A 模型。接着,通过数据验证,使用粒子群算法分阶段对此模型进行参数估计,粗略得到传染病发展的整体趋势。观察结果发现:显性感染人群的变化受较多外界因素影响,而微分方程无法进行模型外生变量的自我调节,导致动力学模型无法达到理想的结果。在深度学习中,时序卷积网络模型能够减少其他因素对预测结果的影响。因此,本文建立 SEIR-A 与 TCN 混合模型来实现每日显性感染者人数的拟合,实验结果证明了该混合模型预测的准确性。

2.1 构建 SEIR-A 模型

“仓室”模型早在 1927 年被首次提出,之后一直广泛应用并得到发展完善。经典传染性模型有 SI 模型、SIR 模型及 SEIR 模型等。本文研究的传染病模型具有隐性感染者居多的特点,且 5 类人群之间存在相应的流动与转换,需要建立符合其发展规律的 SEIR-A 传染病模型。该模型设置了 5 种不同的人群,它们分别为易感人群(S)、潜伏人群(E)、显性感染人群(I)、隐性感染人群(A)以及恢复人群(R)这 5 个不同的仓室,其中重新定义相关人群,详细解释如下。

(1) E:潜伏人群,表示已接触过感染者,但尚未呈现出阳性状态的潜伏人群,同样具有传染性。

(2) A:隐性感染人群,表示尚未出现症状但已呈现出阳性的感染人群,具有传染性。

(3) R:恢复人群,表示通过治疗或者自愈方式获得健康的人群,该阶段的人群具有免疫力且不再具有传染性。

2.1.1 模型假设

通过分析病毒的传播机制(明确易感者、传染源和传播途径)和流行病学相关知识,对疾病传播过程作出必要假设。本文所建立模型基于如下假设:

(1) 不考虑传染期间人群的迁入和迁出、出生与死亡。

(2) 假设潜伏人群和隐性感染人群都具有传染因子,且潜伏人群和显性感染人群具有相同的传染率。

(3) 假设隐性感染患者可以向显性感染患者单向流动,且不可逆。

(4) 假设该类传染病导致死亡的人数极少,可忽

略死亡人群对疾病传播的影响。

2.1.2　确定疾病发生率

已知流行病传染的前提为感染病人与易感者之间发生有效接触,这样才会导致病毒的传播。接触率是指单位时间内一个感染者和易感者之间可能接触的平均次数。由于接触率的形式不同,发生率也会随之改变,常见的形式有双线性发生率、标准发生率及饱和发生率。

通过查阅文献和分析传染病的传播机制,本文选择标准形式的发生率来研究疾病的传播机理,其公式如下:

$$\text{疾病发生率}=\alpha\beta\frac{S(t)}{N(t)}I(t) \tag{1}$$

式(1)中,β 和 N 分别代表疾病传染率和人口总数。

2.1.3　建立微分方程组

假设这一类传染病的传播具有单向不可逆的特点,各类人群的数量随时间而不断变化,t 时刻各人群的状态可分别记作 $S(t)$、$E(t)$、$I(t)$、$A(t)$ 和 $R(t)$,它们之间存在如下恒等关系:

$$S(t)+E(t)+I(t)+R(t)+A(t)=N$$

通过分析疾病传播的现状,可得 SEIR-A 模型的动力学过程如图 1 所示。

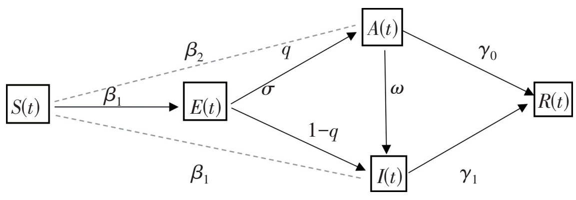


图 1　SEIR-A 模型传播过程图

Fig. 1　Propagation process diagram of SEIR-A model

接着,根据疾病传播规律和发生率的确定,可得到该模型的微分方程组为

$$\begin{cases}\dfrac{\mathrm{d}S(t)}{\mathrm{d}t}=-\dfrac{\alpha\beta_1 I(t)S(t)}{N}-\dfrac{\alpha\beta_1 E(t)S(t)}{N}-\dfrac{\alpha\beta_2 A(t)S(t)}{N}\\[2mm] \dfrac{\mathrm{d}E(t)}{\mathrm{d}t}=\dfrac{\alpha\beta_1 I(t)S(t)}{N}+\dfrac{\alpha\beta_1 E(t)S(t)}{N}+\dfrac{\alpha\beta_2 A(t)S(t)}{N}-\sigma E(t)\\[2mm] \dfrac{\mathrm{d}A(t)}{\mathrm{d}t}=\sigma qE(t)-\omega A(t)-\gamma_0 A(t)\\[2mm] \dfrac{\mathrm{d}I(t)}{\mathrm{d}t}=\sigma(1-q)E(t)+\omega A(t)-\gamma_1 I(t)\\[2mm] \dfrac{\mathrm{d}R(t)}{\mathrm{d}t}=\gamma_0 A(t)+\gamma_1 I(t)\end{cases}$$

其中,模型中的各参数的含义见表 1 所示。

表 1　符号说明

Table 1　Description of symbols

符　号	含　义
N	人口总数
α	一个感染者平均每天接触易感者的人数
β_1	易感者与潜伏人群、显性感染人群之间的传染率
β_2	易感者与隐性感染人群之间的传染率
σ	潜伏期人群的转阳率
ω	隐性感染人群转化为显性感染人群的比率
q	潜伏人群到隐性感染人群的转化率
γ_0	隐性感染人群的移除率
γ_1	显性感染人群的移除率

综上,本文提出改进后的 SEIR-A 模型有如下特点:

(1) 引入具有感染性的隐性感染人群。由于疫苗和易感者免疫力的提高,所以在疾病发展过程中出现了大量没有感染症状的患者,并携带一定的传染因子。

(2) 考虑显性感染人群与隐性感染人群之间的转换率 ω。实际情况下,隐性感染者在医学隔离期间有概率转为显性感染者,两类人群之间存在单向流动。

(3) 分别考虑潜伏人群向隐性感染人群和显性感染人群的转换比率 q、$1-q$。

2.2　构建时间卷积 TCN 模型

在深度学习中,TCN 可以将一维卷积进行变形,建立适用于时序的模型。TCN 建模有两个基本原则,一方面是该模型输入序列的长度决定输出序列的长度,两者保持一致;另一方面指预测结果只与现在和过去有关,与未来无关。为了实现第一点,TCN 模型在架构时采用一维全卷积网络 FCN,这样保证了每个隐藏层的长度与输入层长度相同,再通过设置 padding = 0 使后续层的长度与之前层保持相同。为了实现第二点,可在 TCN 模型中增添因果卷积来保证不能从未来泄露到过去的事实,则模型可表示为

$$\text{TCN}=\text{1D FCN}+\text{因果卷积}$$

接着,为了解决对长序列历史信息的预测问题,采用空洞卷积来实现呈指数级别增大的感受野。对于一维输入序列 $\boldsymbol{X}\in\boldsymbol{R}^n$ 和过滤器 f:$\{0,\cdots,k-1\}\rightarrow\mathbf{R}$,对序列元素 s 的空洞卷积运算 F 可以定义为(d 代表膨胀因子):

$$F(\mathrm{s})=(X_d*f)(s)=\sum_{i=0}^{k-1}f(i)\cdot X_{s-d\cdot i}$$

为进一步解决由深度增长而产生的退化问题,提出残差模块概念。其中 F 的输出被添加到该模块的输入 $\boldsymbol{X}$ 中(σ 为激活函数):

$$o=\sigma(\boldsymbol{X}+F(\boldsymbol{X}))$$

在残差块内,TCN 模型使用了整流线性单元(ReLU),这样不仅扩张了因果卷积,而且存在非线性,并将权重归一化应用于卷积滤波器,实现归一化过程。此外,在每个扩张卷积之后添加了空间 Dropout,目的就是为了正则化:在每个训练步骤中,将整个通道归零。

2.3 构建 SEIR-A 和 TCN 混合模型

本文利用 SEIR-A 模型和 TCN 模型的预测结果进行建模,采用二元线性回归方式进行拟合,结合两者特点组成一种混合模型。线性回归模型主要研究自变量和因变量之间的关系。首先假设 SEIR-A 模型和 TCN 模型存在线性相关性,再通过拟合直线的方式找到两者之间的内部关系,使用该混合模型进行预测的步骤如下:

(1) SEIR-A 模型预测。根据官方数据,利用 SEIR-A 模型分别对起始期、发展期及衰减期进行多阶段建模,在参数估计过程中,采用粒子群算法进行优化,使其适应度函数达到最小值,最终整理得到完整的预测结果 X_1。

(2) TCN 模型预测。对原始数据进行滞后处理来构建适合时序建模的数据集,通过调整不同滑动窗口及设置重要参数的方法来训练模型,并在测试集上进行误差评估,根据不同预测效果选择均方根误差最小的 TCN 模型,其预测结果为 X_2。

(3) SEIR-A 与 TCN 模型混合预测。在得到两个模型的预测结果 X_1 和 X_2 之后,将其作为二元线性回归模型的两个自变量,也就是模型的输入值,用真实数据作为输出值来建立回归模型,其方程可以表示为

$$\hat{y}=[k_1\quad k_2]\begin{bmatrix}X_1\\X_2\end{bmatrix}+b \tag{2}$$

式(2)中,$\hat{y}$ 表示混合模型的预测结果,b 表示偏移量,k_1 和 k_2 表示矩阵系数。

2.4 模型评估指标

本文使用到的模型评估指标一共有 4 个,它们分别是平均绝对误差 Y_{MAE}、均方根误差 Y_{RMSE}、决定系数 R^2 和平均误差 Y_{Bias},将其作为度量该传染病拟合效果的评估体系,可对不同模型进行性能分析,其中定义 Y_{Bias}:

$$Y_{\mathrm{Bias}}=\frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t-\hat{y}_t}{y_t}\right| \tag{3}$$

式(3)中:y_t 表示真实值,$\hat{y}_t$ 表示预测值,n 为样本数。其中 R^2 可以反映模型能否被解释,其数值越接近 1,表明性能越好;其他 3 个指标均在不同角度分析了真实值与预测值之间的偏差,其数值越小越好。

优化思想是参数估计的中心部分。粒子群算法不仅是基于群体的全局随机搜索算法,而且具有智能的特点。在该算法中,可以把每个优化问题的可能解看作搜索空间中的"粒子",将每一个由优化函数决定的适应度值赋给所有的粒子们,同时每一个粒子都具有一个确定飞行方向和距离的速度。接着粒子们在解空间进行最优化搜索,PSO 算法具体步骤如下:

(1) 第一步设置粒子群规模、惯性因子、加速因子以及迭代次数。

(2) 第二步是随机初始化步骤,这也是建模过程的前提基础。一般的做法是初始化粒子的初速度向量,并且假设当前位置就是个体的最优位置,这样减少了复杂度,为后面算法寻找最优位置做铺垫。

(3) 第三步更新每个粒子的速度(约束粒子速度); 更新每个粒子的位置;在 $k+1$ 次迭代 d 维更新位置。

(4) 第四步是模型的中心部分,通过适应度函数的计算结果来更新迭代,方法就是更新每个粒子的个体历史最优解,同时为避免陷入局部最优化,应该设置适当的粒子数量来控制全局最优化。

(5) 第五步,若满足条件,则停止搜索并输出结果;否则回到第三步。

3 模型的验证与评估

3.1 数据来源

为了验证模型的合理性,本文确定研究对象为国内一类含有大量隐性感染者的传染病,且满足模型假设,数据范围为 2022-02-11—2022-06-10,相关数据来源于国家卫生健康委员会官方网站。

3.2 基于粒子群算法验证 SEIR-A 模型

3.2.1 建立多阶段传染病模型

根据真实数据,观察该时间范围内传染病发展现

状(图 2)。其中,现存隐性感染人群和累计恢复人群符合常规变化规律,而现存显性感染人群的发展受外界因素影响较大,具有明显的波动(现存隐性感染人数=累计隐性感染人数-累计隐性感染者自愈人数-累计隐性转显性的感染人数)。

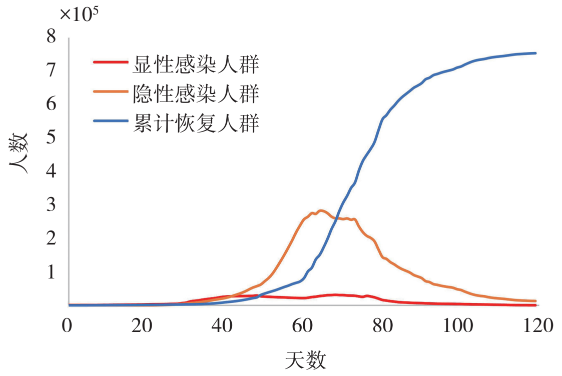


图 2　传染病发展趋势图

Fig. 2　Trend of infectious disease development

由图 2 可以看出:该传染病出现了大量隐性感染患者的现象,表明该疾病传播机制符合本文建立的 SEIR-A 模型,可以进一步用来验证该模型的合理性。由于隐性感染者没有临床表现,导致在生活中无法及时发现,但却在疾病传播过程中起到了关键作用,所以将隐性患者加入感染者中,可以使模型更好地符合实际疾病的发展。现将现存显性感染人群和现存隐性感染人群结合为现有感染人群(现有感染人群=现存显性感染人群+现存隐性感染人群),并以该人群为依据将传染病分为 3 个阶段:2 月 11 日—3 月 2 日(起始期)、3 月 3 日—4 月 17 日(发展期)、4 月 18 日—6 月 10 日(衰减期)。根据不同时期采取不同干预措施,引入不同的参数,建立多阶段传染病模型(图 3)。

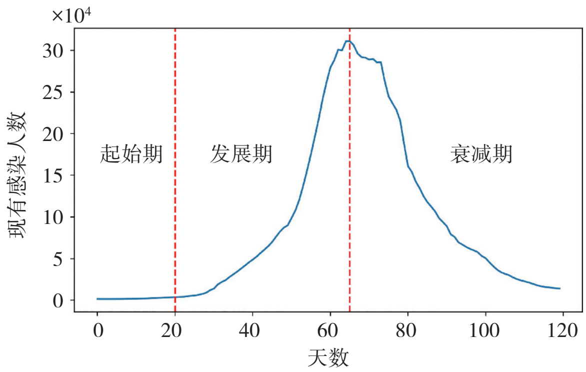


图 3　传染病分段示意图

Fig. 3　Schematic diagram of infectious diseases

3.2.2　Logistic 模型估计易感人群

首先,确定 SEIR-A 模型中各类人群的初始值。一方面,每个感染者可以接触到的人数是有限的,而且一旦被追踪到就会被隔离或者治疗,另一方面,易感者也会通过戴口罩等方式尽量避免与感染者接触。将易感者初值设定为中国总人口是不合理的,在本文中设定 $S(0)$ 是未知量,需要对其进行估计。而潜伏人群会根据易感人群的估计结果,通过解离散化微分方程组的方式获得估计值。此外,显性感染人群数目 $I(0)$、隐性感染人群数目 $A(0)$ 和累计恢复人群数目 $R(0)$ 的初始值由官方数据获得。

人口的增长情况可以用经典的 Logistic 模型来拟合,之前也有学者提出用该模型研究传染病的增长规律,如文献[6]中采用该模型估计得到了易感人群的初始值。由此受到启发,决定采用累计感染人群的数据来估计易感人群,其累计感染人数满足式(4):

$$L(\mathrm{t})=\frac{K}{\left(\frac{K}{L_0}-1\right)\mathrm{e}^{-rt}+1} \tag{4}$$

假定易感者初始值与最大累计感染者在数值上差别不大,式(4)中:L_0 是累计感染人群的初始值;K 的含义是环境容量,这里表示易感人群的初始值,表示传染病模型中的最大累计感染人数;r 表示固定增长率;t 表示时间;$L(t)$ 的含义为随时间发生变化的累计感染人数。拟合结果中决定系数大于 0.95,模型合理。

由上述分析可得易感人群的初始值,则该阶段模型中所有人群的初始值见表 2,总人口数 N=793 060。

表 2　起始期模型初始值

Table 2　Initial values of initial model

人群数目	$S(t)$	$E(t)$	$I(t)$	$A(t)$	$R(t)$
初值	791 246	205	1 436	82	91
方法	估计	计算	真实数据	真实数据	真实数据

显然,易感人群和传染病模型增长趋势呈相反趋势,因此,将 Logistic 拟合模型得到的数值通过降序变化作为易感人群的估计值。同理得到第二阶段和第三阶段的初始人群数量。

3.2.3　粒子群算法估计模型参数值

SEIR-A 模型一共有 8 个参数,其中某些具有现实意义的参数可以根据文献资料获得,如潜伏期的转阳率及易感人群的接触率,分别设置为 0.2 和 14.7;此

外,还可以通过已获得的真实数据计算得到相关参数,后采用取平均值方法来确定该参数,这些可计算的参数分别为 γ_0、γ_1 和 ω。

$$\gamma_0=\frac{\text{累计自愈人数}}{\text{累计隐性感染人数}}$$

$$\gamma_1=\frac{\text{累计治愈人数}}{\text{累计显性感染人数}}$$

$$\omega=\frac{\text{累计隐性转为显性感染人数}}{\text{累计隐性感染人数}}$$

在使用粒子群算法估计 SEIR-A 模型中的其他参数时,粒子代表的就是其他剩余参数,粒子的速度范围就是各个参数的取值范围,其中设置传染率(β_1 和 β_2)的范围为[0,0.5],其余参数在[0,1]中取值,其适应度函数如下所示:

$$f=\sum_{i=1}^{T}(\hat{I}_i-I_i)^2+(\hat{A}_i-A_i)^2+(\hat{R}_i-R_i)^2$$

通过 MATLAB 软件中的 particleswarm( ) 函数来求解该算法,函数采用的是自适应的领域模式,搜索前期使用领域模式,如果适应度开始停滞时,粒子群搜索会从领域模式向全局模式转换。设置学习因子为 1.49,修改粒子数量为 200,其最大的迭代次数为 200。通过不断调试可以发现:在确定数值的 5 个参数中,需要在 γ_0 和 γ_1 中选择一个参数作为已知参数,另一个参数作为未知数,通过求解粒子群算法来获得。

结果显示:第一阶段选择 γ_0 作为已知参数,γ_1 作为未知参数来进行模型的拟合,模型效果如图 4 所示。

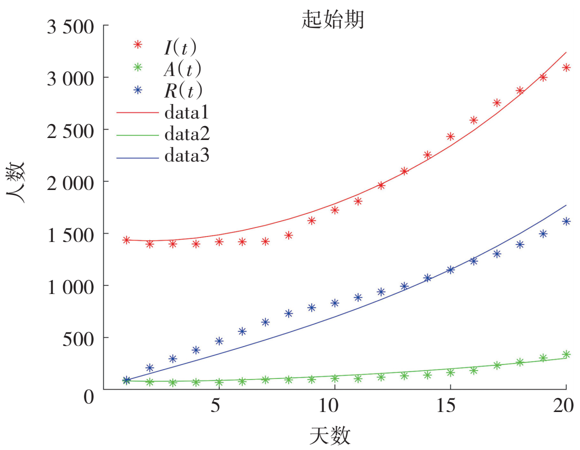
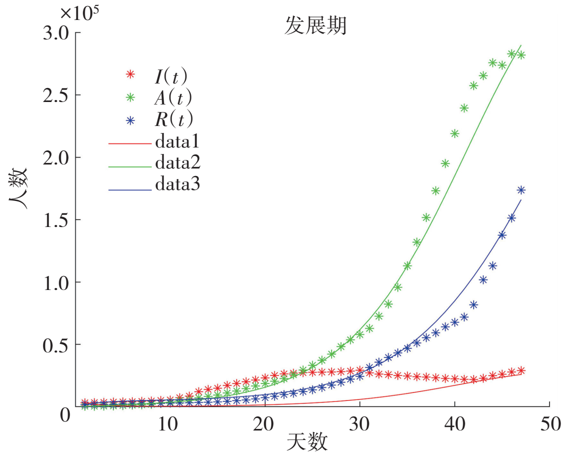


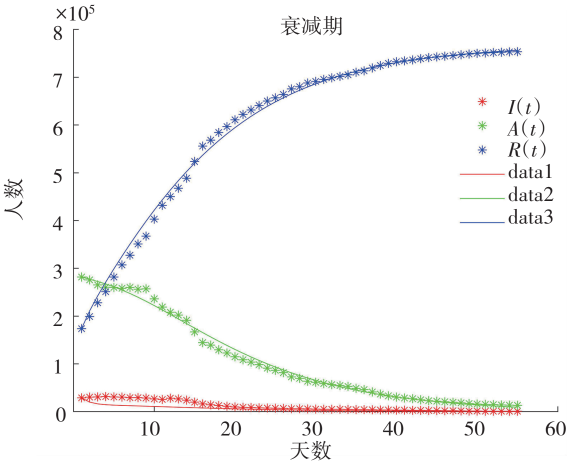
图 4 SEIR-A 模型第一阶段拟合

Fig. 4 First stage fitting of SEIR-A model

接着,第二、三阶段中选择 γ_1 作为已知参数,γ_0 作为未知参数来进行模型的拟合,模型效果如图 5 所示。



(a) 第二阶段



(b) 第三阶段

图 5 SEIR-A 模型第二、三阶段拟合

Fig. 5 Second and third stage fitting of SEIR-A model

经过不断修改调整,得到上述的拟合效果,结果表明该模型具有可行性,可以较好地预测该传染病发展的大致趋势。最终确定 SEIR-A 模型的参数见表 3。

表 3 多阶段 SEIR-A 模型参数估计结果

Table 3 Parameter estimation results of multi-stage SEIR-A model

	起始期	发展期	衰减期	来 源
α	14	14	7	文献
β_1	0.001 6	0.003 0	0.07	估计
β_2	0.104 1	0.023 3	0.008	估计
σ	0.2	0.2	0.2	文献
ω	0.112 6	0.034 8	0.035 4	计算
q	0.178 2	0.881 1	0.865 9	估计
γ_0	0.039 3	0.005	0.054 5	计算/估计
γ_1	0.041 4	0.478 3	0.922 9	计算/估计

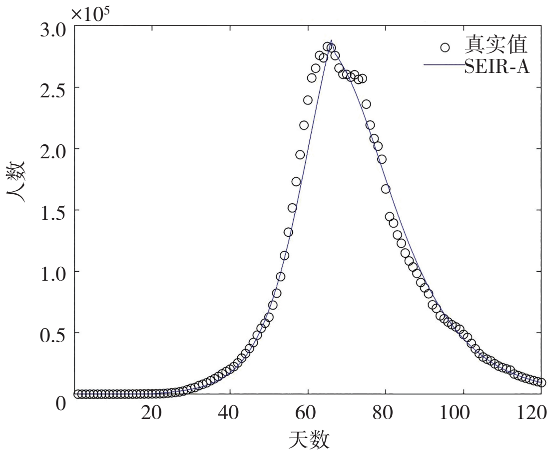
3.2.4 SEIR-A 模型的验证

为了验证模型,分别整理了 3 个阶段的 SEIR-A 模型对现存显性感染人群、现存隐性感染人群以及累计恢复人群的预测结果,并通过评估指标来描述其合理性。其中隐性感染人群和累计恢复人群的拟合结果见表 4 和图 6 所示,模型预测数值与真实数据相差不大,总体预测效果比较好,决定系数都大于 0.95,说明参数估计结果较为准确,反映了 SEIR-A 模型能够科学地模拟传染病的发展规律。

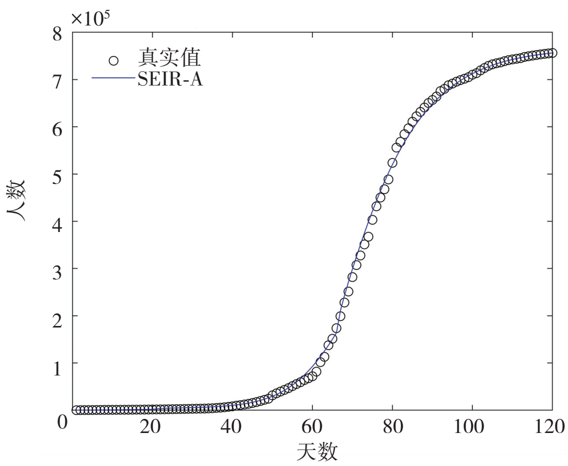
表 4　$A(t)$ 和 $R(t)$ 预测结果评估

Table 4　Evaluation of $A(t)$ and $R(t)$ forecast results

指　标	现存隐性感染人群 $A(t)$	累计恢复人群 $R(t)$
R^2	0.987 0	0.989 9
Y_{Bias}	0.144 1	0.185 3



(a) 现存隐性感染人群



(b) 累计恢复人群

图 6　SEIR-A 模型对现存隐性感染人群和累计恢复人群的拟合

Fig. 6　The fitting of SEIR-A model to recessive infection population and cumulative recovery population

但是,由于现存显性感染人群 $I(t)$ 的增减受不可控因素的影响较大,所以传统动力学模型无法准确预测该人群的动态变化(图 7),导致偏差达到 $Y_{\mathrm{Bias}} = 0.501\ 5$,需要进一步研究分析。

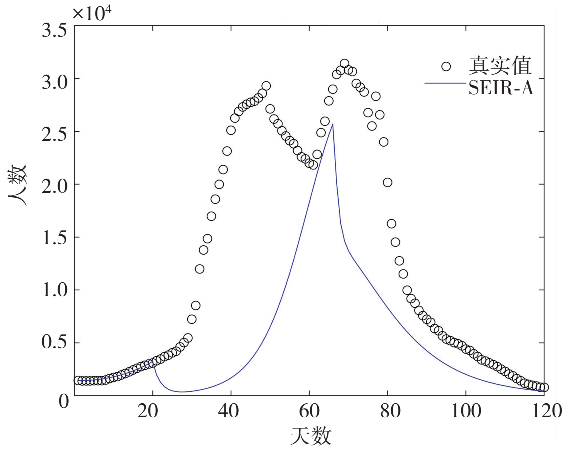


图 7　SEIR-A 模型对现存显性感染人群的拟合

Fig. 7　Fitting of SEIR-A model to existing dominant infection population

3.3 时间卷积 TCN 模型的验证

显然,SEIR-A 模型能够刻画出传染病的总体变化趋势,保证了模型的正确性。但由于不可控因素的影响使预测结果与实际情况存在一定的误差,结合统计学知识,尝试建立关于时间序列的卷积神经网络模型 TCN 来预测现存显性感染人群的变化,并利用 Python 软件得到对每日现存显性感染人数的预测结果。

在建模过程中,TCN 模型把每日现存的显性感染人数作为非线性时间序列,预测过程可以看作是从输入病例空间到输出病例空间的非线性映射,再通过设置合适的滑动窗口,将每日现存显性感染人数时间序列分成输入序列和输出序列,然后使用 TCN 网络挖掘学习输入序列前后之间的相关性,通过全连接网络变换残差块输出序列维度,使其与输出序列维度相同,接着求预测值与真实值的均方根误差,并利用反向传播原理不断更新梯度,最终得到具有预测现存显性感染人群能力的 TCN 模型。

3.3.1 数据归一化

归一化是一种无量纲处理手段,是简化计算和缩小量值的有效办法。在进行不平稳数据训练之前,需要将样本进行归一化处理,让数据缩小至[0,1]范围之内,其计算公式为(式中 $\tilde{X}$ 是归一化后的结果,X 是原始数据集):

$$\tilde{X}=\frac{X-\mathrm{Min}(X)}{\mathrm{Max}(X)-\mathrm{Min}(X)}$$

3.3.2　构建数据集

对于单维时序预测,需要采用滞后处理的方法来修改维度,从而构造出适合 TCN 模型的数据集。设 T 天的每日现存显性感染人数为 $X=(x_1,x_2,\cdots,x_{T-1},x_T)$,$T$ 就是统计数据的天数,假设使用前 T-1 天的每日现存显性感染人群预测第 T 天的显性感染人数,则可以通过设置适合的滞后天数来对 T 天的每日显性感染人数进行划分,将其扩展成为标准的矩阵形式。假设 x_1—x_M 作为第 1 个样本,$(x_1,x_2,\cdots,x_{M-1})$ 为输入数据,x_M 为预测第 M 天的显性感染人数,以此类推,x_2—x_{M+1} 作为第 2 个样本,$X=(x_2,x_3,\cdots,x_{M-1},x_M)$ 为输入的显性感染病例数,x_{M+1} 为预测值,这样就可以得到以下样本矩阵:

$$\boldsymbol{A}=\begin{bmatrix} x_1 & x_2 & \cdots & x_M \\ x_2 & x_3 & \cdots & x_{M+1} \\ \vdots & \vdots & \vdots & \vdots \\ \cdots & \cdots & \cdots & x_T \end{bmatrix}$$

在矩阵 $\boldsymbol{A}$ 中,每一行代表的是一个样本,最后一列代表的是显性感染人群的待预测值。本文设置训练集的占比为 0.8,将该矩阵前 91 个样本作为训练集输入到 TCN 模型中进行参数训练,剩余 23 个样本作为测试集做预测结果的分析验证。

3.3.3　用现存显性感染人群验证 TCN 模型

TCN 模型中输入序列的维数也是滞后处理的天数 M,该参数为待调节参数;本文 TCN 模型采用一维因果卷积,设置 padding 方式是‘causal’; Adam 优化器的选择可以更新步长,它综合考虑了梯度的一阶矩估计和二阶矩估计,提高了计算效率且占用内存较少;最后,为了减少计算量并且避免梯度消失,在残差板块使用 activation=‘relu’作为激活函数,其表达形式为

$$f(x)=\max(0,x)$$

此外,所含残差模块层数、卷积核的个数和维度,以及其他的一些参数都会影响 TCN 模型的训练效率和拟合精度,需要通过不断地调试使模型达到最佳预测性能。假设 M 的取值可能为 2、6、10、14,观察图 8 中 Y_{RMSE} 的变化,由此图可以看出:当 $M=6$ 时得到最小的误差,则确定该模型的输入序列维度。其他重要参数的设置见表 5 所示。

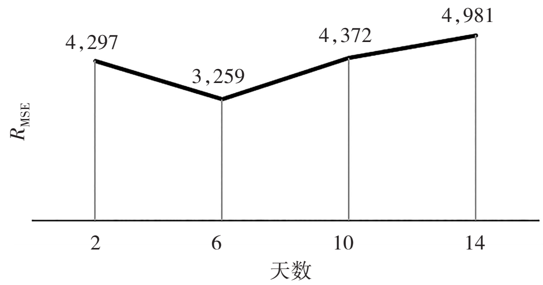


图 8　不同滞后天数下的 R_{MSE} 数值

Fig. 8　R_{MSE} values under different lag days

表 5　TCN 模型中几个重要的参数值

Table 5　Several important parameter values in TCN model

参数	含义	数值
num_hour	滞后处理后时间序列的维数,即输入序列的维数	6
nb_filters	在卷积层中使用的过滤器数量	64
kernel_size	每个卷积层中使用的内核大小	2
batch_size	一次训练所抓取的数据样本数量	64
epoch	将所有训练样本训练几次的过程	200

根据上述参数建立 TCN 模型,预测每日现存显性感染者人数,拟合效果如图 9 所示。$R^2=0.9005$,说明 TCN 模型可以用来预测该传染病中的现存显性感染人群。

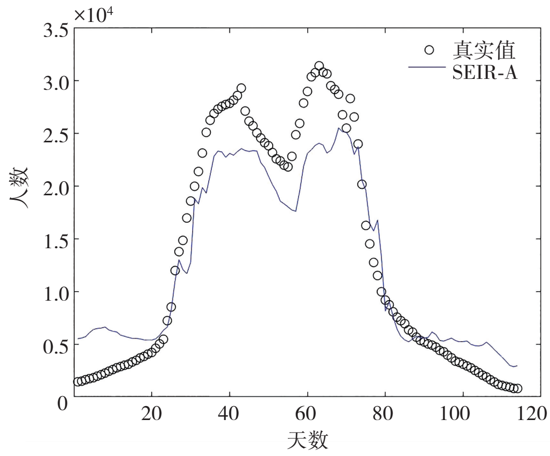


图 9　TCN 模型拟合现存显性感染人群 $I(t)$

Fig. 9　TCN model fitting existing confirmed population $I(t)$

3.4　SEIR-A 与 TCN 混合模型的验证

在 TCN 模型中确定的滞后天数为 6,所以在构建混合模型时选择数据的起始时间为 2022 年 2 月 17 日。分别整理两个独立模型的预测结果作为输入值,每日现存显性感染人群的数据为目标值,利用 Python 软件实现线性回归过程,结果如表 6 所示。其 P 值都小于 0.05,则实验结果具有统计学意义。

表 6　线性回归模型结果

Table 6　Results of linear regression model

	coef	std err	$P>\|t\|$	[0.025	0.975]
const	−3 528.059 6	384.991	0.000	−4 290.945	−2 765.175
x_1	0.119 0	0.044	0.008	0.032	0.206
x_2	1.275 7	0.034	0.000	1.208	1.343

最后，选择近年来被广泛应用于时间序列预测领域的 LSTM 模型作为对比基准，并建立 SEIR-A 与 LSTM 的混合模型。观察图 10 中 5 个模型预测值与真实值之间的拟合图像，发现本文提出的 SEIR-A 与 TCN 混合模型更加贴合现存显性感染人群的变化趋势。

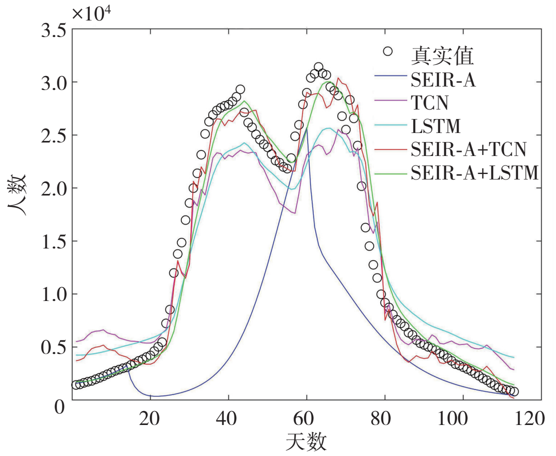


图 10　拟合现存显性感染人群 $I(t)$ 模型对比图

Fig. 10　Fitting the comparison diagram of $I(t)$ model of existing infected population

在数据上，分别计算出 5 种模型预测值与真实值之间的 Y_{MAE}、Y_{RMSE}、R^2 和 Y_{Bias}，发现 SEIR-A 与 TCN 的混合模型和 SEIR-A 与 LSTM 的混合模型相差不大，前者结果略优于后者，决定系数达到 0.9611，且 Y_{MAE}、Y_{RMSE}、Y_{Bias} 分别为 1 628、2 144、0.289 1，都是该对比情况下的最小值。

表 7　显性感染人群 $I(t)$ 拟合结果对比

Table 7　Comparison of $I(t)$ fitting results in dominant infected population

	Y_{MAE}	Y_{RMSE}	R^2	Y_{Bias}
SEIR-A 模型	7 430	10 916	−0.071 6	0.501 5
TCN 模型	2 899	3 459	0.900 5	0.636 2
LSTM 模型	3 146	3 481	0.812	0.600 4
SEIR-A 与 LSTM 混合模型	1 639	2 300	0.955 16	0.317 8
SEIR-A 与 TCN 混合模型	1 628	2 144	0.961 1	0.289 1

综上所有对比实验可以得出：SEIR-A 与 TCN 混合模型对现存显性感染人群的拟合效果相比于 SEIR-A 模型、TCN 模型、LSTM 模型、SEIR-A 与 LSTM 混合模型来说，其拟合效果更为精确，说明该混合模型具有很强的数据拟合能力和时序建模能力。

4　结论与展望

4.1　结　论

本文以传统动力学 SEIR 模型为基础，在原有疾病传播过程中加入具有传染性的潜伏者和隐形传播者，用于表述一类含感染可能性较大的传染病。利用数据进行完整时间段的拟合，得到如下结论：

（1）引入具有不同感染因子的潜伏者和隐性传播者使动力学模型得到完善，利用现存隐性感染人群数量和累计恢复人群数量验证改进后的 SEIR-A 模型，证明该模型合理。

（2）引入时间卷积网络 TCN 模型提高现存显性感染人群的拟合精度，并利用线性回归结合改进后的 SEIR-A 模型得到混合模型，发现该模型在拟合结果上取得最优。

（3）经典动力学和统计学模型的巧妙结合，表明本文提出的 SEIR-A 与 TCN 混合模型不仅体现出了疾病感染的传播机理，而且提高了原有模型的拟合精度。

4.2　展　望

为了对传染病传播进行准确地时序建模，本文提出了一种基于 SEIR-A 与 TCN 混合模型拟合的方法，并利用一类含隐性感染人群传染病的真实情况，验证了该模型的合理性。其 SEIR-A 模型是由经典动力学模型 SEIR 改进而来，在传统定义的基础之上引入了隐性感染人群，并理性增添了相关因子的影响，模型验证过程中采用多阶段拟合的方式，得到了 3 个阶段不同的参数值，这样能有效减少整体预测所产生的误差。结果表明：该模型只适用于对现存隐性感染人群和累计恢复人群进行估计，对现存显性感染人群的拟合有所欠缺。所以，采用时间卷积网络的方式建立 TCN 模型，通过不断调试参数，实现了对显性感染人群更高精度的预测。最后，将数学模型与深度学习利用线性关系结合起来，通过对比可知：该混合模型能够更为准确地分析复杂状态下现存显性感染人群的变化趋势，进一步验证了模型的有效性以及对复杂真实情况的适应性。

但是，本文尚未考虑隔离人群对传染病的影响，后续可以在此基础上完善 SEIR-A 模型，得到更加贴合现

实情况的动力学模型。

参考文献(References):

[1] 李峰. 具有无症状感染的传染病动力学建模分析及应用[D]. 太原: 中北大学, 2021.
LI Feng. Dynamic modeling analysis and application of infectious diseases with asymptomatic infection[D]. Taiyuan: North University of China, 2021.

[2] TANG Y L, XIAO D M, ZHANG W N, et al. Dynamics of epidemic models with asymptomatic infection and seasonal succession[J]. Mathematical Biosciences and Engineering, 2017, 314: 1407—1424.

[3] 马千里, 郭鑫, 申一骏. 基于 SIR-B 对无症状感染者在 COVID-19 中的分析与预测[J]. 山西大学学报(自然科学版), 2022, 45(6): 1—10.
MA Qian-li, GUO Xin, SHEN Yi-jun. Analysis and prediction of asymptomatic infected persons in COVID-19 based on SIR-B[J]. Journal of Shanxi University (Natural Science Edition), 2022, 45(6): 1—10.

[4] 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. Logistic and SEIR combined model predicts the propagation law of novel coronavirus pneumonia[J]. Journal of Xiamen University (Natural Science Edition), 2020, 59(6): 1041—1046.

[5] MALAVIKA B, MARIMUTHU S, JOY M, et al. Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models[J]. Clinical Epidemiology and Global Health, 2021(9): 26—33.

[6] 张晴. 关于新冠肺炎的研究和预测[D]. 济南: 山东大学, 2021.
ZHANG Qing. Research and prediction of epidemic situation in novel coronavirus pneumonia [D]. Jinan: Shandong University, 2021.

[7] 谢小良, 王时雨, 成佳祺, 等. 基于 ARIMA 与 SEIR 模型的突发传染病发病特征与趋势预测[J]. 商学研究, 2022, 29(2): 30—38.
XIE Xiao-liang, WANG Shi-yu, CHENG Jia-qi, et al. Characteristics and trend prediction of sudden infectious diseases based on ARIMA and SEIR models[J]. Business Research, 2022, 29 (2): 30—38.

[8] 郭卫云. 基于 SEIR 模型的传染病预测与可视化系统实现[D]. 邯郸: 河北工程大学, 2020.
GUO Wei-yun. Implementation of infectious disease prediction and visualization system based on SEIR model [D]. Handan: Hebei University of Engineering, 2020.

[9] 薛明劲, 黄钊慰, 胡雨迪, 等. 传染病动力学模型研究进展[J]. 预防医学, 2022, 34(1): 53—57.
XUE Ming-jin, HUANG Zhao-wei, HU Yu-di, et al. Research progress of infectious disease dynamic model[J]. Preventive Medicine, 2022, 34 (1): 53—57.

[10] 刘天, 赵泽宇, 姚梦雷, 等. SEIAR 传染病动力学模型的建立及实现[J]. 疾病监测, 2020, 35(10): 934—938.
LIU Tian, ZHAO Ze-yu, YAO Meng-lei, et al. Establishment and realization of SEIAR infectious disease dynamics model[J]. Disease Surveillance, 2020, 35(10): 934—938.

[11] 黄鹏. 基于机器学习的乙类传染病预测模型研究与实现[D]. 成都: 电子科技大学, 2019.
HUANG Peng. Research and implementation of class B infectious disease prediction model based on machine learning [D]. Chengdu: University of Electronic Science and Technology of China, 2019.

[12] 刘彬彬, 于辛雅, 齐龙兴. 无症状病人对传染病传播影响的模型研究[J]. 应用数学学报, 2021, 44(5): 703—721.
LIU Bin-bin, YU Xin-ya, QI Long-xing. Model study on the influence of asymptomatic patients on the spread of infectious diseases[J]. Journal of Applied Mathematics, 2021, 44 (5): 703—721.

[13] MA Y L, LI H J. Stability research of an SEIR model with distinct general contact rates and infectious force in latent and recovered period [J]. Wuhan University Journal of Natural Sciences, 2022, 27(2): 104—114.

[14] CAI Y L. A stochastic SIRS epidemic model with infectious force under intervention strategies[J]. Journal of Differential Equations, 2015, 259(12): 7463—7502.

[15] HIGAZY M. Novel fractional order SIDARTHE mathematical model of the COVID-19 pandemic [J]. Chaos Solitons & Fractals, 2020, 138(1): 110007—110015.

[16] 白玉珍. 具有控制措施的三类传染病模型的动力学分析[D]. 北京: 北京建筑大学, 2021.
BAI Yu-zhen. Dynamic analysis of three infectious disease models with control measures[D]. Beijing: Beijing University of Civil Engineering and Architecture, 2021.

责任编辑:李翠薇