

基于马尔可夫模型的传染病传播预测研究

孟祥佩

(宁波财经学院 基础学院 浙江 宁波 315175)

【摘要】本文借助马尔可夫模型对传染病传播预测进行了研究。首先,我们对传染病的传播过程进行了建模,将其抽象为离散状态的马尔可夫模型。接着,应用贝叶斯参数估计方法推断了马尔可夫模型的转移概率矩阵。最后,借助 Matlab 软件进行了仿真实验,研究结果揭示了疫苗接种和人口流动因素对传染病感染人数的影响。

【关键词】马尔科夫模型; 疫苗接种; 人口流动; 传染病预测

引言

近年来越来越多的传染病毒侵害人类,如 SARS、禽流感、登革热、埃博拉、MERS、肺炎等,这些病毒均可通过空气传播,传染病已不仅仅是医学问题,更是严重的社会问题^[1]。国内外学者对传染病传播预测的研究主要集中在传染病动力学模型和统计模型两种。典型的动力学模型包括 SIR 模型^[2-3]和 SEIR 模型^[4-5]。SIR 模型是一种传染病动力学模型,将人群划分为易感(Susceptible)、感染者(Infected)和康复者(Recovered)三个状态。需要注意的是,SIR 模型是基于简化假设构建的,实际情况中可能存在许多复杂因素。为了更好地描述潜伏期传染病,人们引入了“暴露者(Exposed)”状态并改进了 SIR 模型,得到了 SEIR 模型。SEIR 模型适用于研究具有潜伏期的传染病,能够更好地描述传染病在人群中的传播过程,包括潜伏期、感染期和康复期。通过模拟不同参数值下的 SEIR 模型,可以帮助预测疫情的发展趋势。需要注意的是,SEIR 模型也是基于简化假设构建的,在实际应用中需要结合具体情境和数据进行调整和改进。

传染病预测的统计模型主要包括马尔科夫模型^[6-8]、线性回归模型^[9-10]和复杂的机器学习模型^[11-12]以及贝叶斯统计模型^[13]。马尔科夫链模型是一种离散状态空间和时间的统计模型,假设与当前的状态有关,与过去的状态无关。在传染病研究中,可以将人群状态划分为不同的类别,通过指定状态转移概率来描述传染病的传播过程。模型在实际应用中需要准确的转移矩阵。线性回归模型可用于探索传染病与各种相关因素之间的关系,从而了解这些因素对传染病传播的影响。线性回归模型操作简便,但需要较大数量的数据。机器学习模型是一种通过训练数据来识别模式并得到预测模型的方法。它基于统计学和算法理论,可

以从大量数据中自动学习,并使用所学知识进行决策或预测。机器学习模型在传染病预测方面具有较高的精度,但同样需要大量数据。

综上所述,传染病动力学模型是微分方程模型,对于传染性较低的传染病风险预测比较精确,但对于传染性较高的传染病传播预测有局限性。马尔科夫模型是一种离散模型,可以弥补这个不足。考虑到传染病传播的时序特点,当前时刻患者人数只与上一时刻患者人数相关,下一时刻患者人数又与当前时刻有着密切关系,这一特征符合马尔科夫随机过程,因此,本文采用马尔科夫模型预测传染病的传播。

1 马尔可夫基本理论

马尔可夫理论的核心是马尔可夫性质,即“未来只与现在有关”,也就是说,给定一个过程当前状态及历史的所有状态,其未来状态仅依赖于当前状态,与历史状态无关。这种性质也被称为无后效性或无记忆性^[14],马尔可夫模型是一种统计模型,用于描述具有马尔可夫属性的随机过程。主要内容包括状态空间、状态转移概率矩阵和初始状态概率。

1.1 状态空间

时间和状态都是离散的马尔科夫过程称为马尔科夫链,简记: $X_n = X(n), n = 0, 1, 2, \dots$, 马尔可夫链是随机变量 $X_1, X_2, \dots$ 的一个数列。这些变量的范围,即他们所有可能取值的集合,被称为“状态空间”。

1.2 转移矩阵及初始概率

在马尔可夫模型中,系统可以在每一步从一个状态转移到另一个状态,也可以保持当前状态不变。这种状态的改变被称为转移。由于状态转移具有随机性,不同的状态改变存在一定的率,这个概率被称为转移概率。 $\pi_i(n) = P(X_n = i)$ 表示在时刻 n 处于状态 i 的状态概率。 $P_{ij} = P(X_{n+1} = j | X_n = i)$

表示在时刻 n 处于状态 i , 在下一时刻 $n+1$ 处于状态 j 的概率。也称为一步转移概率矩阵, 简称转移概率或转移概率矩阵记为 $A = (P_{ij})$ 。系统初始时刻 t_0 的状态概率向量为 $\pi(0) = (\pi_1(0), \pi_2(0), \dots, \pi_k(0))$, 为初始状态概率, 则后续传染病状况的预测为: $\pi(1) = \pi(0)A, \pi(2) = \pi(1)A = \pi(0)A^2, \dots, \pi(k+1) = \pi(0)A^k$ 。如果已知系统初始时刻的状态概率向量和状态转移矩阵 A , 就可以计算出任意时刻系统处于各个状态的概率。

2 贝叶斯估计理论

2.1 后验概率

贝叶斯估计是一种利用贝叶斯定理进行参数估计的方法。它基于贝叶斯定理, 通过将先验知识与观测数据结合, 更新对参数的估计。我们假设参数 θ 是随机变量, 根据观测数据 X 可以得到 θ 的后验概率分布 $P(\theta|X)$ 。其中, $P(\theta)$ 为 θ 的先验概率分布, 表示我们对 θ 的初始认知; $P(X|\theta)$ 为给定 θ 条件下观测数据的似然函数, 表示在已知 θ 的情况下, 观测数据出现的可能性。根据贝叶斯定理, 我们可以得到 θ 的后验概率分布:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}, \quad (1)$$

其中, $P(X)$ 为已知的数据信息, 贝叶斯估计首先要确定先验概率分布 $P(\theta)$ 和似然函数 $P(X|\theta)$ 。先验概率分布可以基于历史数据或专家判断来确定。似然函数则是描述参数与观测数据之间关系的模型。贝叶斯估计能够利用先验信息对参数进行估计, 这对于数据量有限或不完整的情况下尤为重要。通过引入先验知识, 贝叶斯估计可以更准确地估计未知参数, 并有效地处理小样本问题。其次, 贝叶斯估计具有很大的灵活性。用户可以根据具体情况选择不同的先验分布, 并通过后验分布来反映参数的不确定性, 通过后验分布, 我们可以了解参数的不确定程度, 并获得置信区间等有关统计量, 从而更好地理解 and 解释结果。

2.2 转移概率的估计

本文利用贝叶斯估计的方法根据传染病感染、康复、死亡、未感染四个状态的历史数据估计各个状态的转移概率矩阵。首先确定参数的先验概率分布, 本文假设每个转移概率都服从 $\text{Beta}(\alpha, \beta)$ 分布, 其中 α 和 β 是超参数。接下来, 根据历史传染病数据, 根据 (1) 计算每个转移概率的后验概率分布。假设有

n 个样本, 其中第 i 个样本表示在状态 i 转移到状态 j 的次数, 那么第 i 行的转移概率矩阵的元素可以表示为

$$P(j|i) = (\eta + \sum X(i, j)) / (n + m\eta), \quad (2)$$

其中 η 是先验中的超参数, $\sum X(i, j)$ 表示第 i 行中所有转移到状态 j 的次数, m 表示转移概率矩阵的列数。

3 仿真研究

考虑隔离管控、人口流动因素疫情的影响。分别对比隔离管控、人口流动前后病毒感染人数的变化。设置初始状态人口分别为 $A_0 = [1000, 10, 0, 9000]$ (单位: 万人), 根据 2022 年 1 月至 2022 年 12 月四个状态的历史数据, 利用贝叶斯理论估计各个状态的转移概率, 利用 Matlab (2018a 版本) 进行仿真得到以下结果。

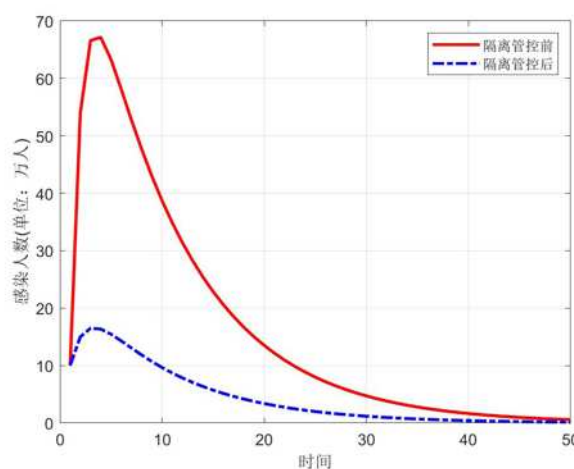


图 1. 隔离管控前后感染人数对比曲线

从图 1 中可以看到隔离管控后的感染人数明显低于隔离管控前的感染人数, 这表明隔离管控对控制传染病的扩散起着非常重要的作用。这是因为隔离管控措施可以有效地阻断病毒在人群中的传播链, 可以限制人们出行、聚集和交往的范围, 减少接触和交叉感染的机会, 从而降低感染人数, 这也是传染病初期最有效的措施。

从图 2 可以看到人口流动前的感染人数低于人口流动后的感染人数, 因为人口流动往往会导致感染人数增加的风险。人口流动使得无症状感染者或轻微症状患者有机会将病毒传播到其他从而导致感染人数的增加。在人口流动前, 感染人数通常相对较低。这是因为在没有大规模人口流动的情况下, 病毒传播的

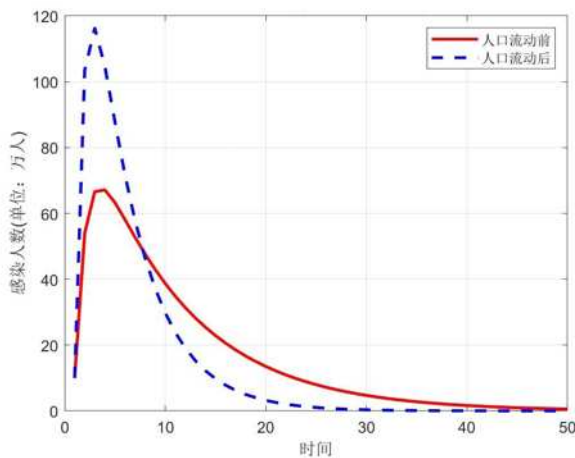


图 2. 人口流动前后感染人数对比曲线

机会相对较少。通过采取及时的隔离控制措施，可以有效遏制病毒传播，从而保持感染人数相对较低。然而，一旦人口开始流动，例如节假日、旅游季节、商务出差等情况下，人们可能从一个地区到另一个地区，这就增加了感染传播的风险。因此，控制人口流动并配合相应的隔离管控措施是减少传染病传播的重要手段。

结论

本文利用马尔可夫模型进行了传染病传播风险的预测。通过对传染病感染人数的预测建模，应用贝叶斯参数估计方法推断了马尔可夫链的转移概率矩阵。根据研究结果揭示了疫苗接种和人口迁移等因素对传染病传播的影响特征。疫苗接种和隔离防控对传染病感染人数的控制是积极的。它不仅可以直接预防感染，减少感染者的数量，从而对整个社会的公共卫生健康产生积极的影响。因此，传染病的控制和预防工作中，推广和普及疫苗接种以及人口流动管理和监测是非常重要的和有效的措施。这些研究结果为公共卫生政策制定者提供了重要的参考，并为进一步研究传染病传播提供了新的思路和方法。

参考文献：

- [1] Kermack, W. O. , Mckendrick, A. G .A Contribution to the Mathematical Theory of Epidemics [J]. Proceedings of the Royal Society A Mathematical Physical & Engineering Sciences, 1927, 115 (772): 700–721.
- [2] Kuznetsov, Yu. A., and C. Piccardi. “Bifurcation analysis of periodic SEIR and SIR epidemic models.”

Journal of Mathematical Biology. 1994,32. (2): 109–121.

- [3] 郭晓霞, 孙树林. 混合随机 SIR 传染病模型的动力学分析. [J]. 系统科学与数, 2022,42 (4): 992–1010.

- [4] Schwartz I B, ,and Smith H L. Infinite subharmonic bifurcation in an SEIR epidemic model.[J]. Journal of mathematical biology, 1983, 18(3): 233–253.

- [5] 曹盛力, 冯沛华, 时朋朋, 修正 SEIR 传染病动力学模型应用于湖北省 2019 冠状病毒病 (COVID-19) 疫情预测和评估 [J]. 浙江大学学报 (医学版), 2020,49 (02): 178–184.

- [6]Greenhalgh D, Liang Y, Mao X. Modelling the effect of telegraph noise in the SIRS epidemic model using Markovian switching[J]. Physica A: Statistical Mechanics and its Applications, 2016, 462.

- [7] 邓甦, 李晓毅. 马尔科夫链在呼吸道传染病预测中的应用 [J]. 中国卫生统计, 2010, (6) 2: 615–616.

- [8] 夏梓桐, 杨春雨, 韩七星. 具有 Markov 切换的非线性随机 SIQS 传染病模型的动力学行为 [J]. 黑龙江大学自然科学学报, 2022, 39(1): 42–48.

- [9] 夏婉玉. 基于多元线性回归方法的疫情监测系统研究 [D]. 湖北: 武汉工程大学, 2022.

- [10] 沈世敬, 崔晓鸣, 曹务春. 传染病预测模型的应用及研究进展 [J]. 中华医院感染学杂志, 2023,33 (16): 2550–2554.

- [11] 曲宗希, 沙勇忠, 李雨桐. 基于灰狼优化与多机器学习的重大传染病集合预测研究—以 COVID-19 疫情为例 [J]. 数据分析与知识发现, 2022,6(8):122–133.

- [12] 董帅兵, 王丽萍, 张业武, 等. 基于机器学习的全国布鲁氏菌病重复报告分析方法研究 [J]. 公共卫生与预防医学, 2022,33(5):29–31.

- [13] 李田, 代雨岑, 李佳蔚, 等. 贝叶斯时空模型在传染病数据时空分析中的研究进展 [J]. 现代预防医学, 2022,49(16):2908–2912.

- [14] 刘次华, 随机过程, 武汉, 华中科技大学出版社, 2003,96–105.

项目：浙江省哲学社会科学规划课题，“基于多元异构大数据的传染病传播风险预测研究”，项目编号：21NDQN290YB。