

基于 SIR-YN 模型的疫情发展趋势预测

2021 年 3 月 14 日

摘要

自 2019 年年末新型冠状病毒首次被发现以后，一场新冠肺炎疫情便迅速爆发并蔓延至全球，不但造成了大量的死亡，而且对世界各国政治、经济等多个方面也都产生了重大的影响。时至今日，人类仍然未能彻底战胜疫情。因此，本文将试图根据中国的疫情状况建立数学模型，并以此来预测中国新型冠状病毒肺炎的发展趋势。

首先，我们根据中国的疫情数据，利用微分方程，在 SIR 传染病模型的基础上，充分结合了中国疫情防控现实状况，对该模型进行了优化修正，从而建立了 SIR-YN 传染病模型，并由此对中国的疫情发展做出了预测。然后，我们将预测数据与实际数据进行对比，评估了所建立模型的精准度，并分析了误差的产生原因。最后，我们针对误差的产生原因，提出了改进模型的方向，并综合评价了该模型的优缺点。

关键字：微分方程、SIR-YN 模型

1 问题背景与重述

1.1 问题背景

2019 年新冠肺炎疫情来袭，对人们的生命造成了巨大威胁，而且对世界各国的政治、经济等多个方面都产生了重大的影响，及时控制疫情的蔓延和发展对个体和国家都有重要意义。为了实施有效的防控措施，建立有效的疫情预测模型十分必要。

1.2 问题重述

根据问题的背景，题目要求建立数学模型解决以下任务：

- 1) 根据已有的疫情数据，分析未来疫情的发展趋势
- 2) 与实际数据对比，验证模型的精确度并进行误差分析
- 3) 对疫情防控提出合理建议

2 模型假设

- 1) 考虑到目前已处于疫情控制阶段，且人们自我隔离意识较好，不妨设每个病人的有效日接触率为定值
- 2) 所有人口都为易感染者，不考虑个别免疫体质

- 3) 疑似病例一旦确诊即被隔离，不会再传播给他人
- 4) 隔离人群中一旦被确定未感染，则立刻结束隔离，即恢复易感者身份
- 5) 在考虑隔离者与未隔离者时将确诊感染者和潜伏期患者都定义为感染者

3 符号说明

符号	基本说明
$S(t)$	表示 t 时刻 易感人群 的总人数
$I(t)$	表示 t 时刻 感染人数 的总人数
$R(t)$	表示 t 时刻 退出者（治愈 + 死亡）的总人数
$Y(t)$	表示 t 时刻 疑似者（实际被感染 + 实际未被感染）的总人数
$N(t)$	表示 t 时刻 所有未隔离患者 的总人数
κ	疑似人群中被确定未感染人数占疑似人群总数的比例
λ	隔离者被确诊人数占隔离者总人数的比例
θ	被隔离的人数占未隔离总人数的比例
ω	被确诊且隔离人数占未隔离总人数的比例
ρ	感染者平均每天对任何状态的人的接触率
ξ	退出率（死亡率 + 治愈率）

4 问题分析

- 1) 根据已有数据建立预测模型，根据传染病的基本过程，确定了影响疫情变化趋势的各个参量，基于微分方程模型进行未来疫情发展趋势的求解。
- 2) 将预测数据与实际数据进行比对，利用统计学方法进行模型精确度分析，并结合实际情况进行误差来源的分析从而改进所建立的模型。
- 3) 根据已经建立的模型，寻求对疫情发展影响较大的参量，并分析如何改变这些参量来控制疫情，从而提出切实的建议。

5 模型建立与求解

5.1 各类人群的转化过程

由于本文针对病毒传播的分析是在疫情得到控制后的阶段，并且我国采取的隔离措施非常果断且高效，因此我们不使用病毒潜伏人群的概念，转而使用疑似者和未隔离患者的概念。也就是说，我们将人群分为易感人群 S 、感染人群 I 、退出者 R 、疑似者 Y 、未隔离患者 N 五类。在此基础上，我们结合中国的疫情防控措施分析了各人群之间的转化：

- 1) 现实中，未隔离人群中既存在易感人群，又存在未隔离的已感染者，这些已感染者与其他人的接触是随机的，也即已感染者不仅会接触易感人群，还会接触已感染者。

- 2) 当感染者被确诊后，其接触过的人群（包括已感染者和易感人群）中的一部分会被追踪到并定为疑似者处理，而未被追踪为疑似者的人中有部分可能转化为患者，即未隔离患者。
- 3) 在疑似病例里面，一部分并没有被感染，仍然为易感人群，而另一部分被感染，由疑似病例转为感染者。
- 4) 感染者经过治疗，有一部分康复，另一部分死亡，我们将康复者和死亡者统称为退出者。

图 1: 各类人群转换流程图

- a) 关于易感者的变化，即疑似人群中未被感染的个体数减去未隔离的个体数。相应计算公式如下：

$$\frac{dS(t)}{dt} = \kappa Y(t) - \rho S(t)N(t) \quad (1)$$

- b) 关于感染者的变化，即为疑似人群中确诊的个体数加未隔离患者个体数再减去退出者个体数，其中，退出个体数即已经治愈以及死亡的个体数之和。相应计算公式如下：

$$\frac{dI(t)}{dt} = \omega N(t) + \lambda Y(t) - \xi I(t), \quad \frac{dR(t)}{dt} = \xi I(t) \quad (2)$$

- c) 关于疑似者的变化，即为未隔离人群转为隔离人群（疑似者）的个体数减去隔离人群中被排除（确定未感染）的个体数，再减去隔离人群中被确诊的个体数。相应计算公式如下：

$$\frac{dY(t)}{dt} = \theta N(t)S(t) - \kappa Y(t) - \lambda Y(t) \quad (3)$$

- d) 关于未隔离患者的变化，即为未隔离患者个体数减去被追踪到的未隔离患者的个体数。相应计算公式如下：

$$\frac{dN(t)}{dt} = (1 - \theta)N(t)S(t) - \omega N(t) \quad (4)$$

5.2 初值选取与参数确定

首先，我们选取了 2 月 1 日的疫情数据作为微分方程的初始值：

$$\begin{cases} S(t_0) &= 140005 \times 10^4 \text{ 人}, \\ I(t_0) &= 14052 \text{ 人}, \\ R(t_0) &= 130 \text{ 人}, \\ Y(t_0) &= 4562 \text{ 人}, \\ N(t_0) &= 6336 \text{ 人}, \end{cases}$$

其次，通过查阅相关文献，我们确定了不同时期的各项参数：

$$\begin{cases} \kappa_1 = 0.06, & \kappa_2 = 0.057 \\ \lambda_1 = 0.44, & \lambda_2 = 0.052 \\ \theta_1 = 0.75, & \theta_2 = 0.6 \\ \omega_1 = 0.163, & \omega_2 = 0.4 \\ \rho_1 = 0.45, & \rho_2 = 0.45 \\ \xi_1 = 0.04, & \xi_2 = 0.25 \end{cases}$$

其中，下标 1 表示 2-3 月的参数，下标 2 表示 3-5 的参数

5.3 模型求解

综上所述，可以得到模型：

$$\begin{cases} \frac{dS(t)}{dt} &= \kappa Y(t) - \rho S(t)N(t) \\ \frac{dI(t)}{dt} &= \omega N(t) + \lambda Y(t) - \xi I(t) \\ \frac{dR(t)}{dt} &= \xi I(t) \\ \frac{dY(t)}{dt} &= \theta N(t)S(t) - \kappa Y(t) - \lambda Y(t) \\ \frac{dN(t)}{dt} &= (1 - \theta)N(t)S(t) - \omega N(t) \end{cases}$$

解得：

$$\text{感染总人数 } \mathbf{I} = \mathbf{a}_1 \mathbf{e}^{-\left(\frac{t-b_1}{c_1}\right)^2} + \mathbf{a}_2 \mathbf{e}^{-\left(\frac{t-b_2}{c_2}\right)^2} + \mathbf{a}_3 \mathbf{e}^{-\left(\frac{t-b_3}{c_3}\right)^2} \quad (5)$$

其中，各项参数为：

$$\begin{cases} a_1 = 9549, & a_2 = 45380, & a_3 = 9187 \\ b_1 = 17.75, & b_2 = 12.94, & b_3 = 28.48 \\ c_1 = 2.609, & c_2 = 13.11, & c_3 = 26.48 \end{cases}$$

图 2: 微分方程组预测感染者与实际值对比

从图中我们可以发现我们的模型在整个时间跨度中具有实际情况符合程度极高的预测结果。另外, 当我们只截取 五月 ($t \in [90, 120]$) 的数据进行对比, 我们的模型仍然能比较好地预测感染人数的变化趋势。

图 3: 微分方程、高斯拟合感染者与实际对比图

6 模型分析

6.1 验证模型精确度

计算预测数据与实际数据的相对误差, 公式如下:

$$E = \frac{\sum_{i=1}^n \frac{|x_i - \hat{x}_i|}{x_i}}{n} = 21.74\% \quad (6)$$

考虑到中国人口基数大, 疫情数据相对人口基数较小, 该误差处于可接受的范围, 又根据图中预测图像及实际图像可知, 预测模型与实际情况基本吻合, 说明该模型精确度较高。

6.2 误差分析

- 在第 10 天左右, 实际数值发生了激增, 我们认为这是由于检测方法的优化和检测能力的提高, 一天能检测的规模更大, 因而产生了激增, 这与我们的预测并不冲突。
- 疫情到达峰值之后的一段时间, 实际数值下降的速率相比预测数值下降速率更慢, 我们认为这是由于感染人数持续较多, 增大了医院治疗的压力, 对医疗物资的消耗较严重, 因此下降速率更慢。

6.3 模型评价

• 模型的优点

- (1) 预测范围较广, 短中长期均可以预测
- (2) 考虑到了现实中多种因素对疫情的干扰, 更具实际意义
- (3) 结合了中国此次抗疫所采取的高效有力的防控措施, 区别于传统的传染病模型

• 模型的缺点

- (1) 没有考虑病毒的活性变化及突发事件
- (2) 忽略了个体的体质等各方面的差异

6.4 模型改进方向

- 重新选取微分方程预测的初始值以及时间段
- 改变微分方程的参数, 将其由常数变为随时间 t 变化的函数, 以更好地拟合实际情况

7 疫情防控建议

- 1) 实施居家隔离、封城、公共领域限制等一系列持续地严防严控举措，进而降低疫情传播规模。
- 2) 对疑似感染者集中快速隔离和诊断，进而缩减疫情扩散区域，阻碍疫情多点散花似扩散。
- 3) 加强建设优质且全方位覆盖的公共卫生基础设施以及提高公共卫生治疗水平。