

---

# 遗传算法在武汉新冠肺炎传染病模型方面的应用

王朋 学号:2019000000\*

February 18, 2020

## 1 摘要

2020年1月,湖北省武汉市爆发了名为“新型冠状病毒肺炎”的疫情,这种疾病由新型冠状病毒(2019-nCoV)引起,本次疫情发展迅速,影响较大。为了对疫情的发展做出预测,我们需要对本次疫情建立模型来做出分析,对疫情地地发展有大概地认识后,就能方便社会各界更好地做出决策。由于新型冠状病毒肺炎有潜伏期,并且潜伏期内具有传染性,因此本文采用改进的传染病模型SEIR模型,在这个模型中S表示易感者,E表示潜伏期患者,I表示感染者,R表示移出者(治愈或死亡)。由于模型中地各项参数不确定,我们使用机器学习的方法确定该模型的参数。具体来说,我们采用遗传算法来计算该模型的参数,由于在疫情刚开始的时候未被社会重视,而到1月17日本次疫情被指出可以人传染人,开始受到国家的高度重视,社会各界也开始逐步提高警惕。因此本次疫情刚开始传播时的传染率较低,而随着时间的发展,传染率逐步变低,所以本文的模型中的传染率为一个变量,并且由参数控制并且由遗传算法通过近数十天的数据分析,得到该参数。最后我们的模型可以与到目前为止的数据很好地拟合,通过模型分析

## 2 研究背景

2019年12月,一场名为新型冠状病毒肺炎的瘟疫开始在中国大陆悄悄传播,直到1月24日武汉市封城,这场才受到公众的关注,各省市也立即出台相应的政策尽力抑制本次疫情的发展,国内外社会各界也向疫情区提供大量的支援和帮助。疫情前期由于未被社会所重视,所以传播较快,直到1月17日,钟南山院士宣布本次的新型冠状病毒可以人传人,才被社会各界关注并提高了警惕。本次新冠病毒肺炎全国的确证人数从2020年1月26日的2761人,到2月15日已经增加到68584人,可见本次疫情来势猛烈,日前已经被世界卫生组织定为“国际关注的突发公共卫生事件”。随着社会各界加强防范,抑制了疫情的发展,每日新增患者数已经得到控制。为了对疫情将来的发展做出准确的预测,以方便人们做出正确的判断和决策,需要借助传染病动力学模型来对本次疫情进行分析。由于本次疫情有潜伏期的存在,我们可以采用带有潜伏期的传染病模型SEIR模型,并且使用机器学习的方法确定模型的参数,来帮助来完成模型。

## 3 预备知识

### 3.1 SIR模型与SEIR模型

SIR模型是一种传播模型,是传染病模型中最经典的模型,其中S表示易感者,I表示感染者,R表示移出者。

SIR传染病模型由Kermack与McKendrick在1927年利用用动力学的方法建立。SIR模型将总人口分为以下三类:易感者(susceptibles),其数量记为 $S(t)$ ,表示 $t$ 时刻未染病但有可能被该类疾病传染的人数;染

---

\*王朋 西北工业大学计算机学院研究生

病者(infectives), 其数量记为 $I(t)$ , 表示 $t$ 时刻已被感染成为病人而且具有传染力的人数; 恢复者(recovered), 其数量记为 $R(t)$ , 表示 $t$ 时刻已从染病者中移出的人数。设总人口为 $N(t)$ , 则有

$$N(t) = S(t) + I(t) + R(t) \quad (1)$$

SIR模型的建立基于以下三个假设:

(1) 不考虑人口的出生、死亡、流动等种群动力因素。人口始终保持一个常数, 即 $N(t) \equiv K$ 。

(2) 一个病人一旦与易感者接触就必然具有一定的传染力。假设 $t$ 时刻单位时间内, 一个病人能传染的易感者数目与此环境内易感者总数 $S(t)$ 成正比, 比例系数为 $\beta$ , 从而在 $t$ 时刻单位时间内被所有病人传染的人数为 $\beta S(t)I(t)$ 。

(3)  $t$ 时刻, 单位时间内从染病者中移出的人数与病人数量成正比, 比例系数为 $\gamma$ , 单位时间内移出者的数量为 $\gamma I(t)$ 。

但在实际情况中, 传染病往往会有潜伏期, 在这个期间的病人虽然染病但未表现出症状, 而且往往具有传染性, 比如本次新冠肺炎就是具有潜伏期的。于是为了将SIR模型适用于具有一定潜伏期的传染病, 在传统的SIR模型上添加了潜伏期E, 就形成了SEIR模型。其状态之间的转化如 Figure 1所示:

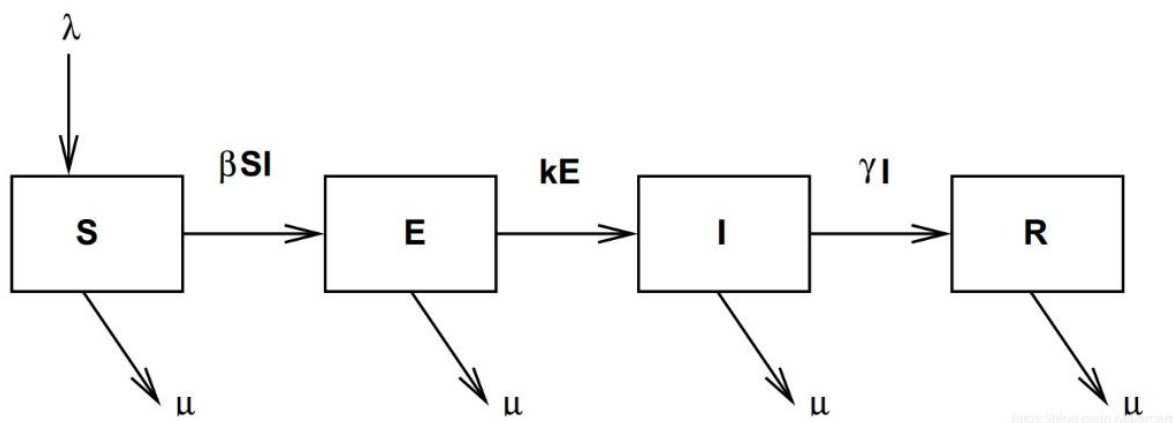


Figure 1: 描述

其中,  $\lambda$  表示易感人群的输入或人口增加,  $\mu$  表示所有人的死亡率, 一旦死亡就从该模型中移除.  $k$  表示从暴露人群到确诊感染者的比率;  $\gamma$  是感染者的治愈率, 感染者一旦被治愈一般不会再患该病。SEIR模型可以由微分方程构建。

$$\frac{ds}{dt} = -\beta S \frac{I}{N} \quad (2)$$

$$\frac{de}{dt} = \beta S \frac{I}{N} - kE \quad (3)$$

$$\frac{dI}{dt} = kE - (\gamma + \mu)I \quad (4)$$

$$\frac{dR}{dt} = \gamma I \quad (5)$$

### 3.2 遗传算法

遗传算法作为进化算法中最成熟的一类算法，现在已经被广泛应用于科研和实际生活中。该算法模拟的是达尔文生物进化论中的自然选择和遗传学机理的生物进化过程，通过模拟自然进化过程进而搜索最优解的方法。

众所周知，达尔文生物进化论中生物进化的单位是种群，而一个种群则由一定数目的个体组成。每个个体的各种性状都由位于染色体上的基因控制。因此，在一开始需要实现从表现型到基因型的映射即编码工作。由于仿照基因编码的工作很复杂，我们往往进行简化，如二进制编码，初代种群产生之后，按照适者生存和优胜劣汰的原理，逐代演化产生出越来越好的近似解，在每一代，根据问题域中个体的适应度大小选择个体，并借助于自然遗传学的遗传算子进行组合交叉和变异，产生出代表新的解集的种群。这个过程将导致种群像自然进化一样的后代种群比前代更加适应于环境，末代种群中的最优个体经过解码，即可作为问题近似最优解。

遗传算法的具体流程为：

- 1、初始化，将进化代数置为0，并且设置最大代数，并且随机地生成初始的种群。
- 2、个体评价，计算每个个体的适应度。
- 3、选择，基于适应度优胜劣汰地从种群中选择若干个体。
- 4、交叉变异，被选择的个体交换基因，基因随机地突变，进而生成新的个体。
- 5、重复步骤2到5，直到到达最大迭代的代数。

## 4 模型建立

### 4.1 传染病模型的建立

由于本次疫情潜伏期地存在，我们采用SEIR模型，而且瘟疫开始流行地前几天，社会并没有对本次疫情很好地重视起来，导致初始几天地传染率较高，随着时间的发展，传染率逐步降低。因此，我们用公式定义了变化的传染率。

在这个公式中，随着时间的增长传染率会越来越低，传染率由两个参数控制。

$$\lambda = 1 - \frac{1}{1 + e^{-\omega(t-d)}} \quad (6)$$

$$\frac{dS}{dt} = \lambda S(I(1 - \beta) + E) \quad (7)$$

$$\frac{dE}{dt} = \lambda S(I(1 - \beta) + E) - \sigma E \quad (8)$$

$$\frac{dI}{dt} = \sigma E - \gamma I \quad (9)$$

$$\frac{dR}{dt} = \gamma I \quad (10)$$