

PDE SIR

王乾宇

2024 年 3 月 6 日

摘要

老师我已完成部分模型搭建的工作（代码见[github](#)），使用的方案二2.2。在将真实数据带入求解事遇到了如下问题（具体见3），应该对年龄和时间分别选择哪个时间单位？比如在这个方程中， a, t 的单位应该一致，但如果时间单位是日、月、年，是否需要调整参数，是否影响结果

$$\begin{cases} \frac{\partial}{\partial a} S(a, t) + \frac{\partial}{\partial t} S(a, t) = -(d(a, t) + (v(a, t) + \theta(a, t))p(a))S(a, t) - S(a, t) \int_0^A \lambda(a, a', t) I(a', t) da' + \delta(a) R(a, t) \\ \frac{\partial}{\partial a} I(a, t) + \frac{\partial}{\partial t} I(a, t) = S(a, t) \int_0^A \lambda(a, a', t) I(a', t) da' - (d(a, t) + \gamma(a)) I(a, t) \\ \frac{\partial}{\partial a} R(a, t) + \frac{\partial}{\partial t} R(a, t) = (v(a, t) + \theta(a, t))p(a)S(a, t) + \gamma(a) I(a, t) - (d(a, t) + \delta(a)) R(a, t) \end{cases} \quad (0.1)$$

目录

1	模型建立	2
2	网络结构问题	3
2.1	存在的问题	3
2.2	解决方案	3
3	参数的问题	4
3.1	时间单位	4
3.2	差值方法的选择	6
4	数据获取	6
4.1	人口年龄构成	6
4.2	分年龄组死亡率	6

1 模型建立

表 1.1: 符号说明

符号	含义
a	年龄
$d(a, t)$	死亡率
$v(a, t)$	疫苗接种率
$\lambda(a, a', t)$	a' 对 a 的传染率
$p(a)$	疫苗保护率
$\theta(a, t)$	补充接种率
$\delta(a)$	免疫丧失率
$B(t)$	出生人口
$M(t)$	母婴传播数
A	人群最大寿命
$\gamma(a)$	恢复率

假设:

1. 由于流行时间较短、数据获取难度较大, $p(a), \delta(a), \gamma(a)$ 只与年龄有关, 不随时间变化
2. 存在母婴传播, 但不存在先天免疫

可能存在的问题:

1. 没有考虑病死率 (所有群体死亡率一致)
2. 没有考虑人口迁入迁出

方程:

$$\begin{cases} \frac{\partial}{\partial a} S(a, t) + \frac{\partial}{\partial t} S(a, t) = - (d(a, t) + (v(a, t) + \theta(a, t))p(a))S(a, t) - S(a, t) \int_0^A \lambda(a, a', t) I(a', t) da' + \delta(a)R(a, t) \\ \frac{\partial}{\partial a} I(a, t) + \frac{\partial}{\partial t} I(a, t) = S(a, t) \int_0^A \lambda(a, a', t) I(a', t) da' - (d(a, t) + \gamma(a))I(a, t) \\ \frac{\partial}{\partial a} R(a, t) + \frac{\partial}{\partial t} R(a, t) = (v(a, t) + \theta(a, t))p(a)S(a, t) + \gamma(a)I(a, t) - (d(a, t) + \delta(a))R(a, t) \end{cases} \quad (1.1)$$

边界条件:

$$\begin{cases} S(0, t) = B(t) - M(t) \\ I(0, t) = M(t) \\ R(0, t) = 0 \\ S(a, 0) = S_0(a) \\ I(a, 0) = I_0(a) \\ R(a, 0) = R_0(a) \end{cases} \quad (1.2)$$

2 网络结构问题

2.1 存在的问题

在之前 ode 的 SEIR 模型 [1] 复现中存在无法同时优化两个网络 (如图2.1), 而导致求解困难, 达不到理想效果的问题。当时的解决方案是把两个网络合并为一个网络。即将

$$\begin{aligned} NN_1(t) &= \lambda \\ NN_2(\lambda, t) &= X \end{aligned} \quad (2.1)$$

改为

$$X, \lambda = NN(t) \quad (2.2)$$

其中 X 为各仓室人口数量, λ 为传染率, 两个网络输入变量相同都是 t , 可以合并。而在 PDE 模型(1.1)中

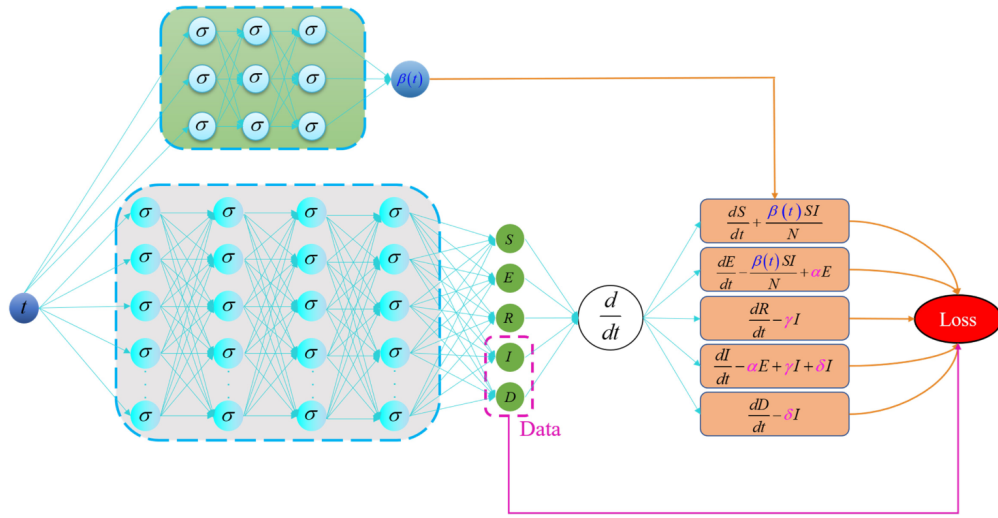


图 2.1: ODE 情况下理论上的网络结构

若采用两个网络则

$$\begin{aligned} NN_1(a, a', t) &= \lambda(a, a', t) \\ NN_2(a, t, \lambda) &= X \end{aligned} \quad (2.3)$$

两个网络的输入变量不同, 由以下解决方案。

2.2 解决方案

- 一. 将(2.3)中的 $NN_2(a, t, \lambda)$ 增添一个输入变量 a' 变为 $NN_2(a, a', t, \lambda)$
- 二. 化简(2.3)中的 $\int_0^A \lambda(a, a', t) I(a', t) da'$ 部分, 由积分中值定理

$$\int_0^A \lambda(a, a', t) I(a', t) da' = \lambda(a, \epsilon, t) \int_0^A I(a', t) da', \epsilon \in (0, A) \quad (2.4)$$

每一对 a, t 对应一个 ϵ , 存在 $\hat{\lambda}(a, t)$

$$\hat{\lambda}(a, t) = \lambda(a, \epsilon(a, t), t) \quad (2.5)$$

可以将(2.3)改为

$$\begin{aligned} NN_1(a, t) &= \hat{\lambda}(a, t) \\ NN_2(a, t, \lambda) &= X \end{aligned} \quad (2.6)$$

其中 $\hat{\lambda}(a, t)$ 的实际含义为年龄为 a 的易感群体在 t 时刻的易感程度。

三. 分布训练两个模型（预训练 + LoRA 微调）

在之前 ode 的 SEIR 模型 [1] 中，损失由数据损失和 ODE 损失两部分组成：

$$Loss = Loss_{data} + Loss_{ODE} \quad (2.7)$$

而在神经网络(2.3)中， $NN_2(a, t)$ 对应 $Loss_{ODE}$ ，可以预先设定不同的 $\lambda(a, a', t)$ ，使用(1.1)生成多组仿真数据预先训练 $NN_2(a, t)$ ，当 $NN_2(a, t)$ 学习到足够多的 PDE 信息后再将其中的参数冻结，用 $NN_1(a, a', t)$ 代替 $\lambda(a, a', t)$ 使用真实数据进行训练。

表 2.1: 总结

方案	优点	不足
一	1. 可以得到每个年龄段对每个年龄段的传染率 $\lambda(a, a', t)$ 2. 保持模型结构不变	1. 增大计算量 2. 现有数据量可能不足
二	1. 大幅减少计算量	1. 只能得到每个年龄段的易感率 $\lambda(a, t)$
三	1. 可以得到每个年龄段对每个年龄段的传染率 $\lambda(a, a', t)$ 2. 在较少数据下表现更好 3. 训练好的 $NN_2(a, t)$ 可以在其他数据集直接使用	1. 增大计算量 2. 如何生成预训练的仿真数据 3. 是否具有可行性

对于这三种解决方案，按照可行性从二 > 一 > 三的顺序进行。

3 参数的问题

3.1 时间单位

对于部分参数，师姐给出了一些文献中的参考值，如（图3.1）为年龄分组的数值，分组的组距为 7 个月、15 个月、12 年、5 年、10 年不等。同时，麻疹数据的时间单位为月，年龄单位为年（如图3.1），对于本文的连续性模型应该采用怎样的时间单位？（月？年？）

如果 a 的单位为年， t 的单位为月的话，(1.1)式应改为：

$$\begin{cases} \frac{\partial}{\partial a} S(a, t) + 12 \frac{\partial}{\partial t} S(a, t) = - (d(a, t) + (v(a, t) + \theta(a, t))p(a))S(a, t) - S(a, t) \int_0^A \lambda(a, a', t) I(a', t) da' + \delta(a)R(a, t) \\ \frac{\partial}{\partial a} I(a, t) + 12 \frac{\partial}{\partial t} I(a, t) = S(a, t) \int_0^A \lambda(a, a', t) I(a', t) da' - (d(a, t) + \gamma(a))I(a, t) \\ \frac{\partial}{\partial a} R(a, t) + 12 \frac{\partial}{\partial t} R(a, t) = (v(a, t) + \theta(a, t))p(a)S(a, t) + \gamma(a)I(a, t) - (d(a, t) + \delta(a))R(a, t) \end{cases} \quad (3.1)$$

参数	含义	月龄	数值	数值来源	浙江省2010-2018年麻疹补充接种情况 (θ)				
$q_0(a)$	各年龄组人群初始免疫率	0≤a<7(0-7月)	q=0	文献查找	时间	年龄	补种人数	补种率%	数据来源
		8≤a<24(8-23月)	待估	文献查找	2010.9	8月-4周岁	2421830	97.5	合作项目数据
		24≤a<180(2-14岁)	待估		2011.10	初三、高一	1011917	>95.0	
		180≤a<240(15-19岁)	0.9089		2012.10	初三	445536	>95.0	
		240≤a<360(20-29)	0.9168		2013.10	初三	426722	>95.0	
		360≤a<480(30-39)	0.915		2014.10	初三	444822	>95.0	
		480≤a<600 (40-50)	0.9275		2015.10	初三	405022	>95.0	
$v(a)$	常规接种率	a≥600	0.9226	文献查找	2016.10	初三	425822	>95.0	
		0≤a<8月	0		2017.10	初三	419627	>95.0	
		8≤a<18	0.95		2018.10	初三	461247	>95.0	
		18≤a<24	0.95		总计		6462545	>95.0	
$p(a)$	疫苗保护率	a≥24	0	文献查找					
		0≤a<8月	0						
		8≤a<18月	0.85						
		18≤a<24月	0.9						
γ	恢复率	无	1						

图 3.1: 部分参数值

诊断年	2005											
诊断月	1	2	3	4	5	6	7	8	9	10	11	12
年龄分 总计	242	383	1112	2395	1970	1013	529	267	124	53	62	153
0-7月	14	37	82	253	238	118	87	43	22	3	7	26
8-17月	19	42	100	302	310	186	88	71	23	12	22	37
18-23月	3	4	9	23	13	7	10	5	2	1	0	0
2岁-	1	12	18	67	68	29	29	15	3	3	1	7
3岁-	4	6	14	38	47	23	14	5	8	0	2	6
4岁-	3	3	10	60	35	22	8	3	3	2	0	1
5岁-	6	8	37	57	52	23	11	2	2	4	2	3
6岁-	4	3	29	60	74	35	11	6	1	1	1	3
7岁-	7	7	37	87	63	27	10	4	4	2	2	0
8岁-	7	6	43	63	56	19	9	3	4	2	3	9
9岁-	5	4	21	49	43	13	4	2	4	1	2	2
10岁-	29	20	96	188	147	85	32	7	5	4	3	26
15岁-	24	26	83	151	81	54	24	9	7	1	4	7
20岁-	43	47	119	174	149	90	40	21	8	5	2	9
25岁-	30	63	147	292	192	98	45	29	10	4	4	7
30岁-	23	53	143	274	206	107	52	20	8	1	1	7
35岁-	13	27	81	174	123	52	38	11	4	6	1	3
40岁-	5	12	36	66	53	19	12	9	4	0	1	0
45岁-	1	3	5	14	13	4	2	2	1	0	0	0
50岁-	0	0	1	3	5	2	3	0	1	1	4	0
55岁-	1	0	1	0	2	0	0	0	0	0	0	0
60岁-	0	0	0	0	0	0	0	0	0	0	0	0
65岁-	0	0	0	0	0	0	0	0	0	0	0	0
70岁-	0	0	0	0	0	0	0	0	0	0	0	0
75岁-	0	0	0	0	0	0	0	0	0	0	0	0
80岁-	0	0	0	0	0	0	0	0	0	0	0	0
>85岁	0	0	0	0	0	0	0	0	0	0	0	0

图 3.2: 麻疹数据

3.2 差值方法的选择

由于能够获取的参数、感染数据都是离散值，而且大部分数据点并没有相互对应。所以需要选择一种差值方法来扩充数据（部分需要二维差值方法）。可以将以下方法都试一遍选择最好的方法：

表 3.1: 插值方法	
方法	效果
牛顿插值	
Hermite 插值	
三次样条插值	

九种常见的二维插值方法

4 数据获取

4.1 人口年龄构成

来源[江苏统计年鉴](#)，根据分年龄段的抽样调查乘以总人数得到各年龄段的总人口。2011 年数据缺失，使用 2010、2012 年的平均值代替。

4.2 分年龄组死亡率

[《江苏省人口平均预期寿命估计及死因差异分析》](#)这篇文章数据由江苏省疾病预防控制中心提供，但在[江苏疾控官网](#)未找到。

References

- [1] Mengqi He et al. “Transmission dynamics informed neural network with application to COVID-19 infections”. In: *Computers in Biology and Medicine* 165 (2023), p. 107431.