

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## 第十三届“认证杯”数学中国

### 数学建模网络挑战赛

### 承 诺 书

我们仔细阅读了第十三届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们接受相应处理结果。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：**32491**

参赛队员（签名）：

队员 1：

王鹤翔

队员 2：

夏晓帆

队员 3：

任文

参赛队教练员（签名）：

肖楠

参赛队伍组别（例如本科组）：本科组

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

---

第十三届“认证杯”数学中国

数学建模网络挑战赛

编号专用页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：

32491

竞赛统一编号（由竞赛组委会送至评委团前编号）：

---

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## 2020 年第十三届“认证杯”数学中国 数学建模网络挑战赛第一阶段论文

题 目 基于对 COVID-19 疫情防控的复工复产研究

关 键 词 因子分析法 GA-BP 神经网络模型 通径分析

引入无症状感染者 SEIR 模型 COVID-19

### 摘 要

**针对问题一**，为了更好地对新冠疫情的高峰进行量化评估，本文延续使用第一阶段所建立的流行病流行程度评估指数  $F$ ，并在选取了此次疫情中比较典型的三个国家进行研究，分别为处于亚洲的中国、处于北美洲的美国以及处于欧洲的德国。基于 GA-BP 神经网络模型对所研究国家的疫情进行预测得到其第二次疫情高峰的预测时间，并对所研究国家的二次高峰的风险进行了量化，基于以上研究给出了针对不同国家的复工复产建议。

**针对问题二**，本文首先综合考虑潜伏期的长度、无症状感染者的比例、无症状感染者的假阴性率、无症状感染者的假阳性率、感染者数量、治愈人数、病死人数等因素，基于因子分析构建起一个能够评价防疫工作及疾病流行的评估指数  $Fa$ 。其次进行通径分析，量化自变量和因变量之间的复杂关系，从而得到对  $Fa$  影响显著因素重要性的排序。最后，分别剔除最重要的三个因素，通过 GA-BP 神经网络模型预测  $Fa$ ，将预测后的结果与原  $Fa$  进行对比分析，从而可以更好地分析因参数准确性对防疫工作和疾病流行造成的影响。通过模型求解，我们得到最重要的三个因素为潜伏期的长度、无症状感染者的假阴性率、无症状感染者的假阳性率，三个因素如果数据偏低，均极有可能会使得防疫工作开展困难加大。

**针对问题三**，本文结合第一阶段构建的基于引入无症状感染者的 SEIR 模型，以及北京市 COVID-19 现有的疫情数据，对北京市未来几个月的无症状感染者等各群体人数进行预测，预测结果显示无症状感染者的数量在不断降低，整体感染者的数量在缓慢上升，但增幅较小。最后根据预测的结果，并结合大型体育赛事的特点，对重启大型体育赛事进行可能性评估，并分阶段给出了赛事重启的时间表。

**针对问题四**，根据前面建模过程中的数据分析以及模型结果，阐述本团队对于顺利重启一些大型体育赛事有关于疫情防控的备忘录。

参赛队号： 32491

所选题目： C 题

参赛密码

(由组委会填写)

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

---

## Abstract

**Aiming at question 1**, in order to better evaluate new crown to quantify the peak of the outbreak, the article continues to use the first stage of epidemic prevalence evaluation index  $F$ , and typically in the outbreak was chose to study the three countries, respectively in the United States and China, in North America Asia Germany in Europe. Based on the GA-BP neural network model, the prediction time of the second peak of the epidemic in the studied country was obtained by predicting the epidemic situation in the studied country, and the risk of the second peak was quantified. Based on the above research, Suggestions for resuming work and production in different countries were given.

**Aiming at question 2**, this article first considers the length of the incubation period, the proportion of asymptomatic infections, false negative rate of asymptomatic infections, asymptomatic patients, the number of cure, the number of false positive rate, infection factors such as deaths, based on factor analysis to build up a evaluation index to evaluate epidemic prevention and disease popular  $Fa$ . Secondly, path analysis was conducted to quantify the complex relationship between independent variables and dependent variables, so as to obtain the ranking of the importance of significant factors influencing  $Fa$ . Finally, the three most important factors were eliminated, and  $Fa$  was predicted by GA-BP neural network model. The predicted results were compared and analyzed with the original  $Fa$ , so as to better analyze the impact of parameter accuracy on epidemic prevention and disease epidemic. By solving the model, we found that the three most important factors were the length of incubation period, the false negative rate of asymptomatic infected persons and the false positive rate of asymptomatic infected persons. If the data of the three factors were low, it would be very likely to make the epidemic prevention work more difficult.

**Aiming at question 3**, this paper combined with the first stage construction of SEIR model based on the introduction of asymptomatic infections, as well as the Beijing COVID - 19 existing epidemic data, asymptomatic infections of Beijing in the coming months the number of each group such as prediction, prediction results show that the number of asymptomatic infections continue to fall in, the number of overall infection in slowly rising, but small. Finally, based on the predicted results and combined with the characteristics of large-scale sports events, the possibility of the restart of large-scale sports events was evaluated, and the schedule of the restart was given by stages.

**Aiming at question 4**, based on the data analysis and model results in the previous modeling process, this paper elaborated the team's memorandum on epidemic prevention and control for the successful restart of some large-scale sports events.

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## 目录

一. 问题的背景与重述.....	1
1.1 问题背景 .....	1
1.2 问题重述 .....	1
二. 问题分析 .....	1
三. 模型假设 .....	2
四. 符号说明 .....	2
五. 数据预处理 .....	2
5.1 异常值剔除.....	2
5.2 缺失值处理.....	2
5.3 数据标准化.....	3
六. 模型的建立与求解.....	3
6.1 问题一模型的建立与求解 .....	3
6.1.1 流行病流行程度评估指数 $F$ 的建立过程 .....	4
6.1.2 基于 GA-BP 神经网络的第二次峰值预测 .....	6
6.1.3 观测国家第二次峰值预测结果 .....	9
6.1.4 复工复产建议 .....	10
6.2 问题二模型的建立与求解 .....	11
6.2.1 防疫工作及疾病流行评估指标的指标选取 .....	11
6.2.2 基于因子分析法的防疫工作及疾病流行评估指数建立 .....	12
6.2.3 基于通径分析的防疫工作及疾病流行因素研究 .....	14
6.2.4 基于去除 LEN 的防疫工作及疾病流行综合预测模型 .....	15
6.2.5 基于去除 FNR 的防疫工作及疾病流行综合预测模型 .....	17
6.2.6 基于去除 FPR 的防疫工作及疾病流行综合预测模型 .....	18
6.3 问题三模型的建立与求解 .....	18
6.3.1 引入无症状感染者的 SEIR 模型.....	19
6.3.2 北京市感染者预测结果.....	20
6.3.3 针对北京市的赛事重启安排 .....	23
6.4 疫情防控备忘录 .....	24
七. 模型的评价 .....	25
八. 模型的应用与推广 .....	25
九. 参考文献 .....	26
附录 .....	26

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

## 一. 问题的背景与重述

### 1.1 问题背景

2020 年初,新型冠状病毒引发的肺炎疫情席卷全球,随着全球新冠疫情拐点的到来,各国都在启动全面复工、复产的计划,但是必须承认这次疫情有出现第二次高峰的风险,第二次高峰一旦出现可能会更加可怕,对于经济的影响可能是致命的。

对一种刚刚出现的、传染迅速的流行病而言,有许多疾病的特征是不易准确测量的。当这些参数的取值不同时,防疫工作应以何种形式开展可能就会出现差异,疾病流行的最终趋势也会有所不同。

我国的无症状感染者数量持续降低,但是并未清零,也有一些无症状感染者未被发现,给大型体育赛事的举办造成影响。对大型体育赛事重启的可能性进行评估,研究其阶段性重启的方法与时间对于推进复工复产,让人们的生活回到正轨是十分有意义的。

### 1.2 问题重述

基于以上问题背景,本文需建立合理的数学模型来解决以下问题:

1. 建立数学模型,选择三个国家进行研究,评估它们出现新冠疫情第二次高峰的风险大小,并给出复工复产的政策性建议,以避免新冠疫情第二次高峰的出现。
2. 建立合理的数学模型,讨论哪些疾病的特征参数是最重要的,如果这些参数不准确,会对防疫工作和疾病流行的过程带来怎样的影响。
3. 结合第一阶段的模型,充分考虑我国的疫情现状,评估重启大型体育赛事(比如中超足球联赛或者 CBA 篮球联赛)的可能性,并给出分阶段(无观众赛事、部分观众赛事、全部观众但要求戴口罩赛事、全面放开赛事)重启的时间表。
4. 为了能够重启一些大型体育赛事,给有关部门写一份有关于疫情防控的备忘录。

## 二. 问题分析

**针对问题一**,首先需要选取合适的风险描述指标,对现在的疫情进行刻画,对之后的疫情风险进行预测。其次要选择不同地区的国家进行分析对比,给出其第二次高峰的风险预测,并对其复工复产的推进给出合理的政策性建议。

**针对问题二**,首先需要综合考虑潜伏期的长度、无症状感染者的比例、无症状感染者的假阴性率、无症状感染者的假阳性率、感染者数量、治愈人数、病死人数等因素,基于因子分析构建起一个能够评价防疫工作及疾病流行的评估指数  $Fa$ 。其次,通过途径分析量化自变量和因变量之间的复杂关系,对  $Fa$  影响显著的因素进行排序。最后将剔除最重要因素后通过 GA-BP 神经网络模型预测出的  $Fa$  与未剔除任何最重要因素时的  $Fa$  进行对比,从而可以更好地分析因参数准确性对防疫工作和疾病流行造成的影响。

**针对问题三**,本文根据第一阶段构建的基于引入无症状感染者的 SEIR 模型,对无症状感染者等各群体人数进行仿真预测。根据北京市 COVID-19 现有的疫情数据以及预测的结果,结合大型体育赛事的特点,对重启大型体育赛事进行可能性评估,并分阶段给出重启的时间表。

**针对问题四**,我们根据问题一、二、三所研究的成果,阐述本团队对于顺利重启一些大型体育赛事有关于疫情防控的备忘录。

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

## 三. 模型假设

1. 假设附件中所给的数据真实有效。
2. 假设所研究地区人口自然出生率与死亡率无剧烈变化。
3. 假设人口是均匀混合的, 因此我们考虑标准发生率。
4. 假设确诊者被完全完全隔离, 不再具有传染性。
5. 假设北京市的医疗资源充足, 未被确诊的感染者以给定的速率进入确诊室。

## 四. 符号说明

符号	说明
$F$	流行病流行程度评估指数
$Fa$	防疫工作及疾病流行评估指数
$S$	易感者
$E$	潜伏者
$I$	感染者
$R$	移出者
$U$	无症状感染者

## 五. 数据预处理

### 5.1 异常值剔除

对所收集的数据采用拉依达准则进行异常值剔除, 对于每一组时间序列数据当偏差大于  $3\sigma$  时, 该数据为异常值。  $\sigma$  的计算公式如下:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad (\text{式 5-1})$$

式中  $x_i$  分别各指标的时间序列数据,  $\bar{x}$  分别为各指标时间序列数据的平均值。  $n$  为各组数据的样本容量。

异常值判断公式为:

$$|x_i - \bar{x}| > 3\sigma \quad (\text{式 5-2})$$

正常值的范围为  $\bar{x} - 3\sigma \leq x_i \leq \bar{x} + 3\sigma$ 。

### 5.2 缺失值处理

对于查找的数据存在缺失的情况, 本文采取的方式是若缺失值较多, 就放弃该因子; 若缺失值较少, 就采用插值的方法补全。

本文选用三次样条插值方法对缺失数据进行补全, 具体方法如下:



# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

设  $y = f(x)$  在点  $x_0, x_1, x_2, \dots, x_n$  的值为  $y_0, y_1, y_2, \dots, y_n$ ，函数  $S(x)$  满足下列条件

(1)  $S(x_i) = f(x_i) = y_i, i = 0, 1, 2, \dots, n$ ;

(2) 在每个子区间  $[x_i, x_{i+1}] (i = 0, 1, 2, \dots, n-1)$  上  $S(x)$  是三次多项式;

(3)  $S(x)$  在  $[a, b]$  上二阶连续可微。

则称  $S(x)$  为函数  $f(x)$  的三次样条插值函数。

## 5.3 数据标准化

由于指标之间存在不同的量纲，数据大小差别较大，数据范围也不尽相同，会增大某些指标的影响程度同时会减弱某些指标的影响程度，对接下来的分析预处理造成较大的影响。所以须将指标进行归一化处理，转化为  $[0, 1]$  之间的数。

常用的数据标准化方法有以下两种：

### 1. min-max 标准化

最小-最大标准化方法是对原始数据的线性变换。对所有量化后的指标进行归一化处理的公式如下：

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (\text{式 5-3})$$

式中  $x_i$  为各指标序列数据， $x_i'$  为归一化后各指标序列数据， $x_{\max}$ 、 $x_{\min}$  分别为各指标原始序列数据的最大值和最小值。此方法优势在于能够提升归一化程度，其缺陷在于当有新数据加入时需重新定义最大值与最小值。

### 2. 均值方差标准化

均值方差标准化是指也称零-均值标准化，经过处理的数据的均值为 0，标准差为 1。转换公式如下：

$$x' = \frac{x - \mu}{\sigma} \quad (\text{式 5-4})$$

式中， $\mu$  为原始数据的均值， $\sigma$  为原始数据的标准差。

本文在建模过程中使用均值方差标准化与 min-max 标准化相结合的方法，对数据进行标准化处理。

## 六. 模型的建立与求解

### 6.1 问题一模型的建立与求解

为了更好地对新冠疫情的高峰进行量化评估，本文延续使用第一阶段所建立的流行病流行程度评估指数  $F$ ，并在选取了此次疫情中比较典型的三个国家进行研究，选取国家状况如表 6-1 所示。



# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

表 6-1 研究国家疫情现状

大洲	国家	累计确诊人数（截至 5.14）
亚洲	中国	84469
北美洲	美国	1456526
欧洲	德国	174894

下面简述流行病流行程度评估指数  $F$  的建立过程。

## 6.1.1 流行病流行程度评估指数 $F$ 的建立过程

### ● 指标选取

为构建流行病流行程度评估指数,本文选取的指标如表 6-2 所示

表 6-2 流行病流行程度评估指数构建

一级指标	二级指标
疫情状况 $x_1$	感染者数量 $x_{11}$
	治愈人数 $x_{12}$
	病死人数 $x_{13}$
人口状况 $x_2$	人口数量 $x_{21}$
	人口密度 $x_{22}$
经济状况 $x_3$	GDP $x_{31}$
医疗条件 $x_4$	医疗卫生机构数量 $x_{41}$
防疫政策 $x_5$	是否强制全民佩戴口罩 $x_{51}$

### ● 基于因子分析法的流行病流行程度评估指数建立

本文以此次新冠肺炎期间所收集到的中国数据为例,展示评估指数建立过程。对所收集到的各地区各二级指标的日数据进行如下处理:

#### (1) KMO 检验和巴特利特球形检验

KMO 检验是 Kaiser, Meyer 和 Olkin 提出的,该检验是对原始变量之间的简单相关系数和偏相关系数的相对大小进行检验,主要应用于多元统计的因子分析。其中, Kaiser 给出一个 KMO 检验标准:  $KMO > 0.9$ , 非常适合;  $0.8 < KMO < 0.9$ , 适合;  $0.7 < KMO < 0.8$ , 一般;  $0.6 < KMO < 0.7$ , 不太适合;  $KMO < 0.5$ , 不适合。

巴特利特球形检验是一种检验各个变量之间相关性程度的检验方法。用于判断变量是否适合用于做因子分析。其对应的  $p$  值小于用户心中的显著性水平 (一般为 0.05), 那么应该拒绝原假设, 认为相关系数不可能是单位阵, 即原始变量之间存在相关性, 适合于作因子分析。相反不合作因子分析。利用 SPSS 对所收集的数据进行以上两种检验, 其结果如表 6-3 示。

表 6-3 KMO 和 Bartlett 的球形度检验结果

KMO 取样适切性量数		0.860
Bartlett 的球形度检验	近似卡方	28968.508
	自由度	21
	显著性	0.000

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

从表 6-3 检验结果来看，KMO 值大于 0.8，说明变量之间存在相关性；其次根据 Bartlett 球形检验的结果，显著性为 0.000 小于 0.05，说明所收集的数据符合因子分析的前提条件。

## （2）碎石图分析

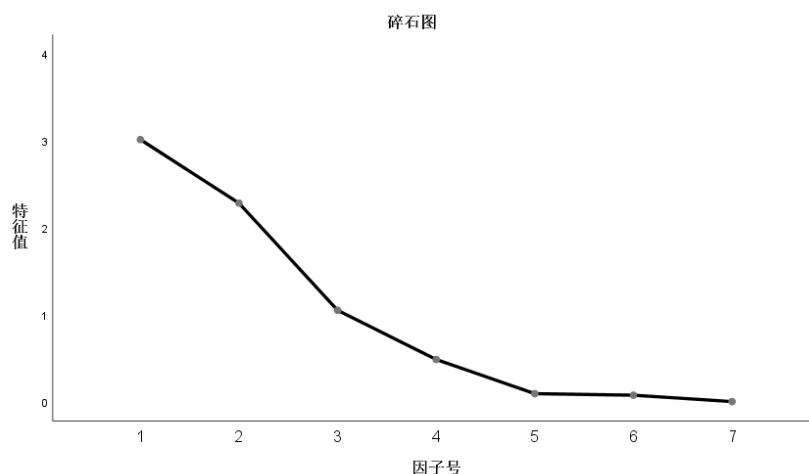


图 6-1 因子分析碎石图

参照因子分析碎石图，各指标数据在因子个数为 3 的时候达到了基本收敛，且累计解释方差基本达到 85% 以上。即可以通过降维后的 3 类因子概括 85% 以上的原始数据信息，具有一定的可靠性和有效性。

## （3）降维分析因子旋转矩阵

旋转后的每个公共因子上的载荷分配更清晰，因而比未旋转时更容易解释各因子的意义。旋转后的因子载荷矩阵如表 6-4 所示，旋转后的因子图如图 6-2 所示。

表 6-4 旋转后的因子载荷矩阵

指标	Factor1	Factor2	Factor3
感染者数量	0.983	0.071	-0.152
治愈人数	0.961	0.058	0.212
病死人数	0.997	0.046	-0.041
人口密度	-0.029	-0.234	0.804
人口数量（万人）	0.050	0.000	0.912
GDP（亿元）	0.059	0.828	-0.012
医疗卫生机构数量	-0.011	0.665	0.020

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

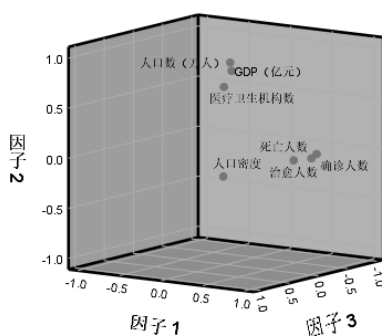


图 6-2 旋转后的因子图

根据因子旋转矩阵与因子旋转图，我们对降维后的三类因子进行解释和定义。

1. factor1 与感染者数量、治愈人数、病死人数关联度最高，我们将其归结为疫情状况指标  $f_1$ ；

2. factor2 与 GDP、医疗卫生机构数量关联度最高，我们将其归结为社会指标  $f_2$ ；

3. factor3 与人口密度、人口密度关联度最高，我们将其归结为人口指标  $f_3$ 。

## ● 流行病流行程度评估指数 $F$ 建立

基于因子分析得到的指标贡献率与旋转后的因子载荷矩阵，我们进一步建立三大因子与流行病流行程度评估指数  $F$ 。各因子指标所占方差百分比与累计贡献率如表 6-5 所示。流行病流行程度评估指数  $F$  如式 6-1 所示。

表 6-5 各因子所占方差百分比与累计贡献率

因子	方差百分比	累计贡献率
1	43.043	43.043
2	32.601	75.644
3	15.011	90.655
4	6.932	97.587
5	1.323	98.910
6	1.071	99.981
7	0.019	100.000

$$\begin{cases} f_1 = 0.983x_1 + 0.961x_2 + 0.997x_3 - 0.029x_4 + 0.05x_5 + 0.059x_6 - 0.011x_7 \\ f_2 = 0.071x_1 + 0.058x_2 + 0.046x_3 + 0.234x_4 + 0.828x_5 + 0.665x_7 \\ f_3 = -0.152x_1 + 0.212x_2 - 0.041x_3 + 0.804x_4 + 0.912x_5 - 0.012x_6 + 0.02x_7 \\ F = 0.475f_1 + 0.360f_2 + 0.165f_3 \end{cases} \quad (\text{式 6-1})$$

式中  $f_1 f_2 f_3$  为降维后的三类因子， $x_i (i=1 \cdots 7)$  为感染者数量、治愈人数、病死人数等指标。

## 6.1.2 基于 GA-BP 神经网络的第二次峰值预测

首先将每日各指标数据集进行训练集与测试集分类，以相邻三天的数据作为训练与

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

验证集合，假设当前时间为  $t$  天，则以  $t-1$  与  $t-2$  两天数据作为训练集；以  $t$  天数据作为训练集。以训练集数据训练模型得到参数后带入测试集，预测之后的各指标观测值，并计算其流行程度评估指数  $F$  观测其下一个峰值的状况。

## ● BP 神经网络原理

BP 神经网络预测模型的实质是根据输入实际情况中所涉及的多维变量进而输出多维所求变量的函数，即能学习和存贮大量的输入-输出模式映射关系，并利用该神经网络函数进行预测，神经网络的演算过程即为调节该函数模型精度的过程，此过程无需事前揭示描述这种映射关系，从而使问题简化且得到有效地解决。BP 神经网络模型的拓扑结构包含：输入层、隐含层和输出层，如图 6-3 所示。

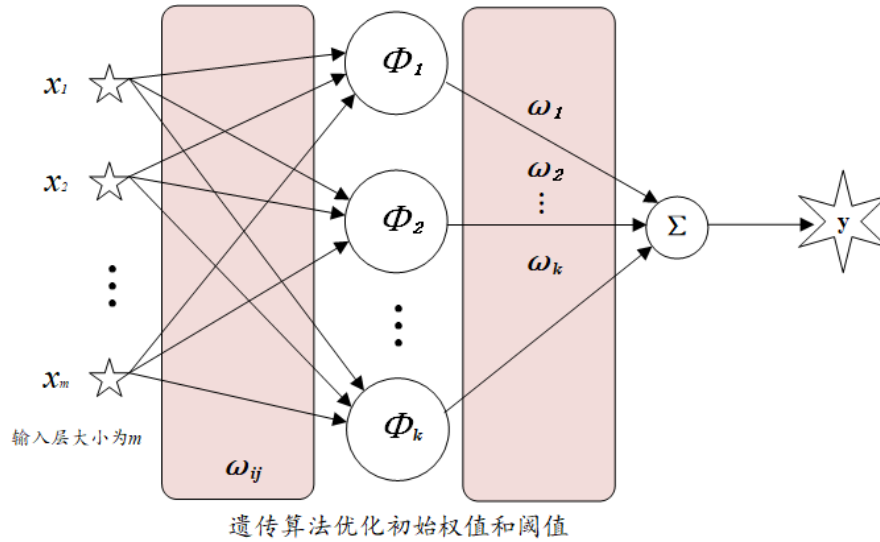


图 6-3 BP 神经网络模型的拓扑结构

## ● BP 神经网络结构

有研究表明三层的神经网络结构具有逼近任意有界连续函数的能力，因此本文选择三层的神经网络进行建模。BP 神经网络输入节点为 2 个，输出节点为 1 个，本文在选取隐层节点个数的问题上参照了以下的经验公式： $l = \sqrt{n + m} + a$  其中， $l$  为隐层节点个数， $n$  为输入层节点个数， $m$  为输出层节点个数， $a$  为  $[1,10]$  之间的常数。根据上式可以计算出神经元个数为 3-12 个。本文选择隐层神经元个数为 5。

输入层到隐层的激活函数采用 Sigmoid 型函数，隐含层到输出层的激活函数采用线性函数 Purelin 函数，可得网络输入与输出的关系如下所示：

$$\hat{y}^k = \sum_{j=1}^r v_j f \left[ \sum_{i=1}^m w_{ij} \cdot P_i + \theta_j \right] (k=1,2,\cdots,N) \quad (\text{式 6-2})$$

其中， $\hat{y}^k$  为神经网络的实际输出值， $y^k$  为期望的输出值， $w_{ij}$  为权值， $\theta_j$  为阈值。

以感染者数量指标的预测为例，所确定的“输入层—隐含层—输出层”的结构示意图如图 6-4 所示。

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

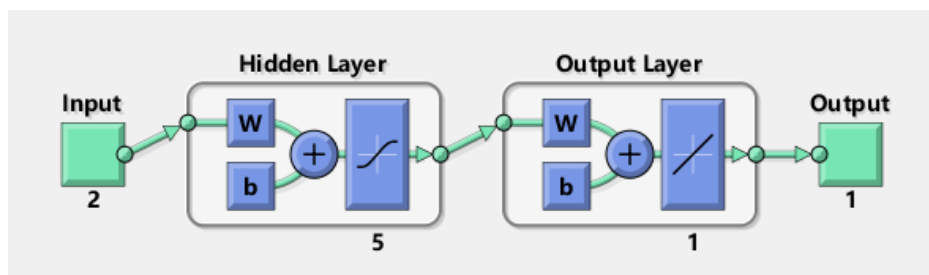


图 6-4 “输入层—隐含层—输出层”的结构示意图

## ● 遗传算法优化 BP 神经网络的过程

遗传算法是一种寻优算法，在人工系统中实现特定目标的优化。遗传算法可通过群体搜索技术，逐个筛选，最终得到最优解或准最优解，将局部和整体有机的结合起来。可用其优化 BP 神经网络预测模型。遗传算法优化的 BP 神经网络结构包括确定个数的变量、层数和算法的复杂程度。利用遗传算法得到 BP 网络结构的最优个体作为权值和阈值，进而利用 BP 神经网络模型的主要功能得到预测值。其改进流程如图 6-5 所示。

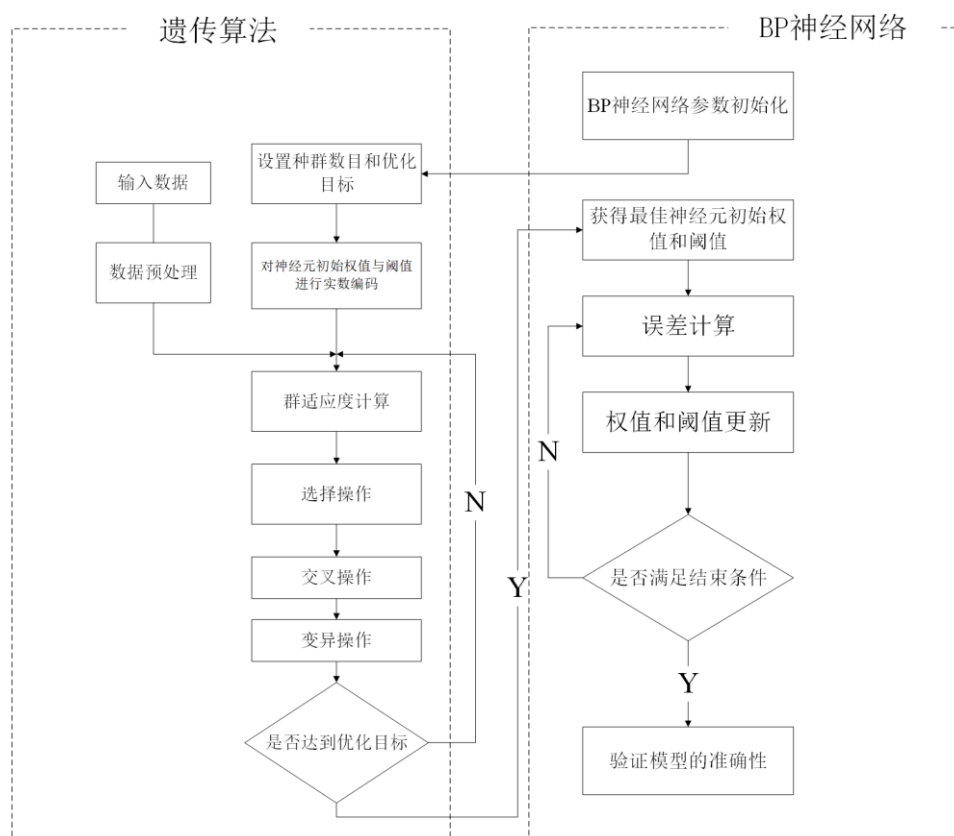


图 6-5 遗传算法改进 BP 神经网络流程示意图

(1) **整体初始化**：将数据进行实数编码。其中，每组数据包含了数据整体的初始权值和阈值，在此过程中，遗传算法将参数初始化，便于后续步骤的进行，输入输出变量已知，此时构成了一个网络清晰、权值和阈值确定的神经网络。

(2) **适应度函数**：由每组数据我们得到初始权值和阈值，然后利用 BP 神经网络对

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

训练集训练,得到预测值,把预测值和实际值之间的误差绝对值作为每组数据的适应值  $F$ ,计算公式如下:

$$F = k \sum_{i=1}^n |y_i - x_i| \quad (\text{式 6-3})$$

式中,  $n$  是网络的输出节点,  $y_i$  是  $BP$  神经网络第  $i$  个节点的期望输出,  $x_i$  为第  $i$  个节点的预测输出,  $k$  是系数。

(3) 选择操作: 利用适应值进行适应度比例法完成操作, 计算公式如下<sup>[2]</sup>:

$$f_i = k / F_i, P_i = \frac{f_i}{\sum_{j=1}^N f_j} \quad (\text{式 6-4})$$

式中,  $F_i$  为个体  $i$  的适应度值,  $k$  为系数,  $N$  为数据组总数。

(4) 变异操作: 选取第  $i$  组数据的第  $j$  个元素  $a_{ij}$  进行操作, 计算公式如下:

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{min})f(g) \\ a_{ij} + (a_{min} - a_{ij})f(g) \end{cases} \quad (\text{式 6-5})$$

式中, 上面的式子限制条件为  $r \geq 0.5$ , 下面的式子的限制条件为  $r < 0.5$ ,  $f(g) = r_2(10g / G_{\max})$ ,  $r_2$  是任意实数,  $r \in [0, 1]$ 。

## ● 优化后 BP 神经网络预测模型中的主要参数

BP 神经网络模型的主要参数包括: 最大训练步数、最大确认失败次数、学习速率、最小性能梯度、训练目标误差。

针对感染者数量指标的预测设置的模型参数如表 6-6 所示。

表 6-6 优化后 BP 神经网络预测模型中的主要参数

最大训练步数	最大确认失败次数	学习速率	最小性能梯度	训练目标误差
1000	5	0.01	1e-6	1e-3

基于遗传算法优化的 BP 神经网络模型的初始化参数有: 迭代次数、种群规模、交叉概率选择、变异概率选择, 如表 6-7 所示。

表 6-7 基于遗传算法优化的 BP 神经网络模型的初始化参数

迭代次数	种群规模	交叉概率选择	变异概率选择
10	30	0.3	0.1

## 6.1.3 观测国家第二次峰值预测结果

以 2020.02.01 为起点, 对所研究国家的所有的指标进行预测, 并计算流行程度评估指数  $F$ , 对第二次高峰进行研究。将所研究国家的流行程度评估指数  $F$  的变化趋势绘制

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

在一张图上，如图 6-6 所示。

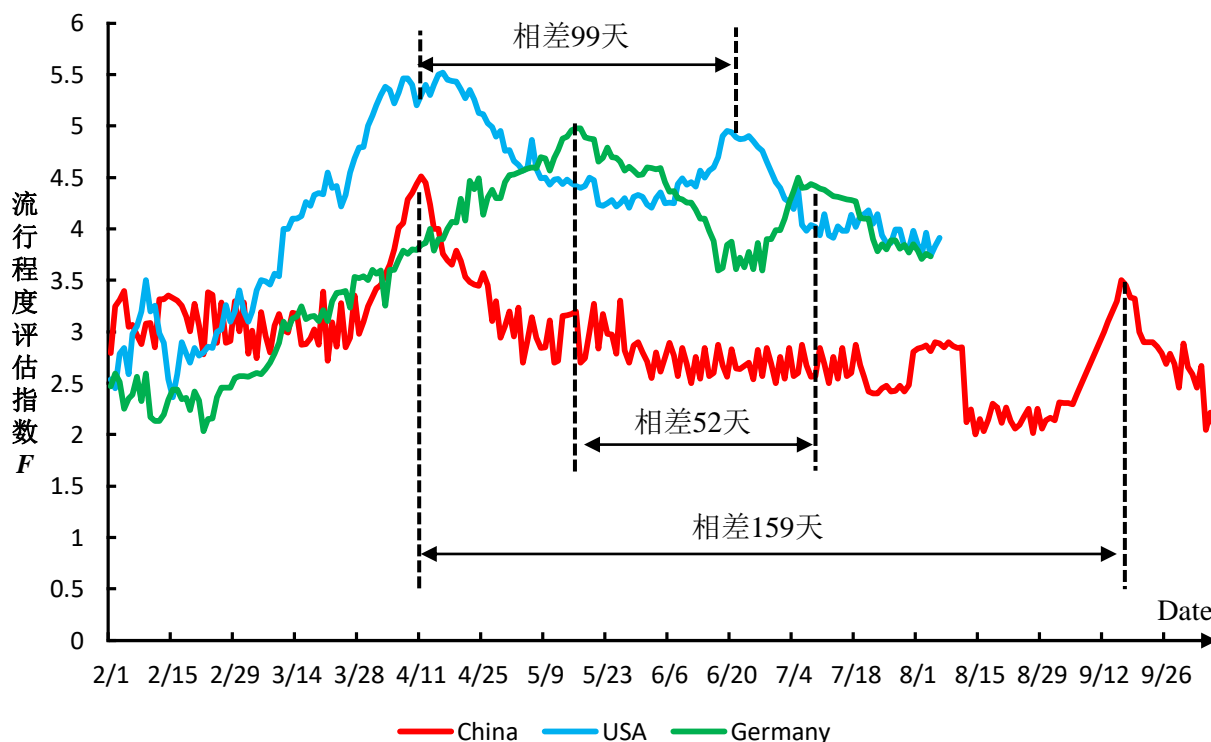


图 6-6 流行程度评估指数  $F$  的变化趋势

所研究国家第一、二次新冠肺炎疫情峰值与出现时间具体情况如表 6-8 所示。

表 6-8 各因子所占方差百分比与累计贡献率

	第一次高峰		第二次高峰	
	峰值	时间	峰值	时间
中国	4.51	4/11	3.46	9/17
美国	5.48	4/6	4.95	6/19
德国	4.98	5/17	4.44	7/8

由上述研究可知，美国两次疫情峰值均高于其他国家，其两次高峰值相差时间低于中国高于美国；德国两次疫情峰值低于美国高于中国，其两次高峰值相差时间最短；中国两次疫情峰值均低于其他国家，两次高峰值相差时间最长。为了量化评估以上三个国家发生第二次疫情高峰的风险，本文定义了如下风险指数  $R$ 。

$$R = \frac{\text{峰值均值}}{\text{峰值时差}} \quad (\text{式 6-6})$$

由上式可知两次峰值均值越大风险越高，两次峰值时差越大风险越低。由此可得德国出现二次疫情峰值的风险最高，美国次之，中国最低。

## 6.1.4 复工复产建议

### ● 针对中国的建议



# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

由上述预测可知，以现有的疫情防控手段，有可能会在9月中旬出现新冠肺炎第二次疫情高峰，出现的原因可能为九月份为学生开学的集中时间，会有大量的大学生跨省人口迁徙，这与中国的人口迁徙特点是非常相关的。也可能与中国的气候特点有关，九月份中国大部进入秋季，人体的免疫能力会有所下降。因此本团队提出如下建议：

(1) 中国需要积极推进疫情防控常态化管理，对于重点地区、重点时间段要严防死守，对于开学集中时间要分批次、分等级、分地区管理复学，切实对高校驻地所在城市做好风险评估工作。

(2) 秋季卫生部门要做好监管工作，提醒人民在换季期添衣，出门佩戴好防护用品，在流感高发季节注意做好与新冠肺炎疫情的区分工作。

## ● 针对美国的建议

美国第一次疫情的峰值偏高，若不采取有效的防控措施极有可能会第二次大规模爆发。爆发的原因可能与其6月份的SAT升学考试等因素有关。建议如下：

(1) 建议美国政府提倡国民少出行、出行佩戴口罩，不要急于推进复工复产，优先支持医疗行业复工复产，保障疫情防控设施与护具的供应能力。

(2) 建议美国有关部门暂缓升学考试与留学生招收工作，保障学生的人身安全，减少新冠肺炎的海外输入。

## ● 针对德国的建议

德国的疫情峰值虽不高但是预测的两次峰值之间的时间差较短，其出现第二次疫情高峰的风险较大，针对德国建议如下：

(1) 建议德国政府实行严格的出入境管理制度，对于欧盟成员国的外来人员也要进行新冠肺炎检测与隔离。

(2) 建议德国政府优先支持医疗行业复工复产，以积极应对新冠肺炎疫情的爆发。

## 6.2 问题二模型的建立与求解

为了更好地研究一些因素对于防疫工作和疾病流行的过程带来的影响，本文基于因子分析法构建了防疫工作及疾病流行评估指数，并通过通径分析量化自变量和因变量之间的复杂关系，从而更加准确地选出对于防疫工作影响较大的因素，更好地开展下一步的防疫工作。

### 6.2.1 防疫工作及疾病流行评估指数的指标选取

为构建防疫工作及疾病流行评估指数，本文选取的指标如表6-9所示。

表 6-9 防疫工作及疾病流行评估指数构建

影响因素		影响因素	
无症状感染状况	潜伏期的长度 LEN	疫情整体状况	感染者数量 NUM
	无症状感染者的比例 PER		治愈人数 REC
	无症状感染者的假阴性率 FNR		病死人数 DIE
	无症状感染者的假阳性率 FPR		

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## 6.2.2 基于因子分析法的防疫工作及疾病流行评估指数建立

### ● KMO 检验和巴特利特球形检验

利用 SPSS 对所收集的指标数据进行检验，检验结果如表 6-10 所示。

表 6-10 KMO 和 Bartlett 的球形度检验结果

KMO 取样适切性量数		0.867
Bartlett 的球形度检验	近似卡方	25867.343
	自由度	21
	显著性	0.000

从表 6-10 检验结果来看，KMO 值大于 0.8，说明变量之间存在相关性；其次根据 Bartlett 球形检验的结果，显著性为 0.000 小于 0.05，说明所收集的数据符合因子分析的前提条件。

### ● 碎石图分析

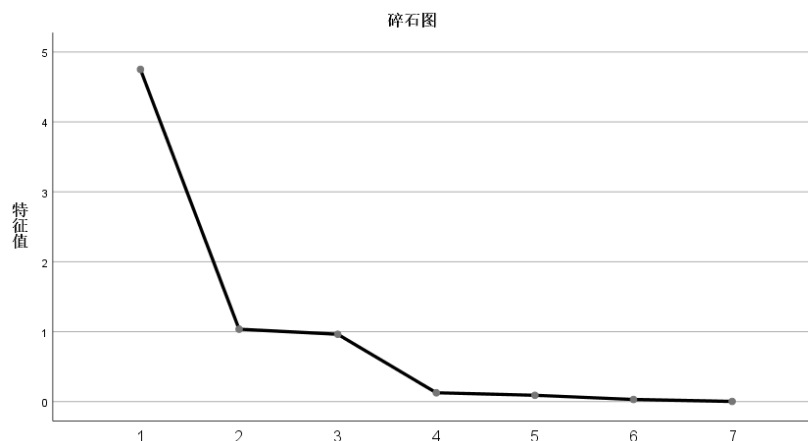


图 6-7 因子分析碎石图

参照因子分析碎石图，各指标数据在因子个数为 2 的时候达到了基本收敛，且累计解释方差基本达到 82% 以上。即可以通过降维后的 2 类因子概括 82% 以上的原始数据信息，具有一定的可靠性和有效性。

### ● 降维分析因子旋转矩阵

旋转后的每个公共因子上的载荷分配更清晰，因而比未旋转时更容易解释各因子的意义。旋转后的因子载荷矩阵如表 6-11 所示，旋转后的因子图如图 6-8 所示。

表 6-11 旋转后的因子载荷矩阵

指标	Factor1	Factor2
LEN	-0.983	-0.921
PER	-0.792	-0.764
FNR	-0.855	-0.752
FPR	-0.786	-0.692
NUM	0.865	0.128
REC	0.632	0.014
DIE	0.842	0.254

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

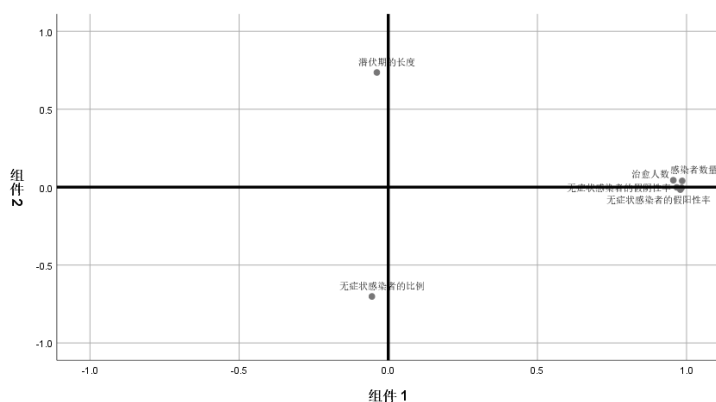


图 6-8 旋转后的空间组件图

根据因子旋转矩阵与因子旋转图，我们对降维后的三类因子进行解释和定义。

(1) factor1 与无症状感染者的假阴性率、无症状感染者的假阳性率、感染者数量、治愈人数、病死人数关联度最高，我们将其归结为疫情状况指标  $fac_1$ ；

(2) factor2 与潜伏期的长度、无症状感染者的比例关联度最高，我们将其归结为社会指标  $fac_2$ ；

## ● 防疫工作及疾病流行评估指数建立

基于因子分析得到的指标贡献率与旋转后的因子载荷矩阵，我们进一步建立两大因子与防疫工作及疾病流行评估指数  $Fa$ 。各因子指标所占方差百分比与累计贡献率如表 6-12 所示。流行病流程度评估指数  $Fa$  如式 6-10 所示。

表 6-12 各因子所占方差百分比与累计贡献率

因子	方差百分比	累计贡献率
1	67.867	67.867
2	14.802	82.669
3	13.759	96.428
4	1.804	98.231
5	1.294	99.526
6	0.435	99.961
7	0.039	100.000

$$fac_1 = -0.983LEN - 0.792PER - 0.855FNR - 0.786FPR + 0.865NUM + 0.632REC + 0.842DIE \quad (\text{式 6-7})$$

$$fac_2 = -0.921LEN - 0.764PER - 0.752FNR - 0.692FPR + 0.128NUM + 0.014REC + 0.254DIE \quad (\text{式 6-8})$$

$$Fa = 0.821fac_1 + 0.179fac_2 \quad (\text{式 6-9})$$

$$Fa = -0.921LEN - 0.764PER - 0.837FNR - 0.769FPR - 0.687NUM + 0.521REC + 0.737DIE \quad (\text{式 6-10})$$

式中  $fac_1$   $fac_2$  为降维后的两类因子。

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## 6.2.3 基于途径分析的防疫工作及疾病流行因素研究

途径分析建立在回归分析的基础上<sup>[1]</sup>,将自变量的途径系数分解成为三部分:直接途径系数、间接途径系数和总途径系数。其中某一自变量对因变量的直接作用效果和间接作用效果分别通过直接途径系数和间接途径系数表现,总途径系数表示各自变量对因变量的综合作用效果,从而量化自变量之间及自变量与因变量的复杂关系。具体方法如下<sup>[1]</sup>:

设自变量为  $v_1, v_2, \dots, v_n$ , 因变量为  $y$ ,  $v_i$  与  $v_j$ 、 $v_i$  与  $y$  之间的简单相关系数分别用  $r_{ij}$  和  $r_{iy}$  表示, 则简单相关系数分解表示为:

$$\begin{cases} P_{1y} + r_{12}P_{2y} + r_{13}P_{3y} + \dots + r_{1m}P_{my} = r_{1y} \\ r_{21}P_{1y} + P_{2y} + r_{23}P_{3y} + \dots + r_{2m}P_{my} = r_{2y} \\ \dots\dots\dots \\ r_{m1}P_{1y} + r_{m2}P_{2y} + r_{m3}P_{3y} + \dots + P_{my} = r_{my} \end{cases} \quad (\text{式 6-11})$$

式中  $P_{iy}$  是直接途径, 即  $v_i$  与  $y$  标准化之后的偏相关系数, 代表  $v_i$  对  $y$  的直接影响效应;  $r_{ij}P_{jy}$  为间接途径, 代表变量  $v_i$  通过  $v_j$  对因变量  $y$  的间接影响效应,  $\sum r_{ij}P_{jy}$  表示  $v_i$  通过其他解释变量对因变量  $y$  的总间接作用。上述方程组的意义是将自变量  $v_i$  与因变量  $y$  的简单相关系数分为  $P_{iy}$  (直接途径效果部分) 与  $\sum_{j \neq i} r_{ij}P_{jy}$  (总间接途径效果部分)。防疫工作及疾病流行影响因素的途径分析结果如表 6-13 所示。

表 6-13 防疫工作及疾病流行影响因素的途径分析系数表

自变量	简单相关系数	直接途径系数	间接途径系数							
			合计	通过 LEN	通过 PER	通过 FNR	通过 FPR	通过 NUM	通过 REC	通过 DIE
LEN	-0.968	-0.783	-0.937	—	0.564	-0.328	-0.684	-0.422	-0.321	0.254
PER	-0.866	-0.654	-0.708	-0.328	—	-0.324	-0.447	0.321	-0.453	0.523
FNR	-0.957	-0.732	-0.903	-0.453	-0.567	—	-0.123	0.486	0.234	-0.487
FPR	-0.876	-0.684	-0.800	-0.503	-0.466	-0.122	—	0.404	0.103	-0.216
NUM	-0.843	-0.576	-0.638	-0.334	-0.302	0.318	0.143	—	-0.231	-0.232
REC	0.765	0.488	-0.604	-0.233	-0.254	-0.546	0.343	0.032	—	0.054
DIE	0.784	0.543	-0.615	-0.289	-0.318	-0.124	0.435	-0.221	-0.109	—

潜伏期的长度、无症状感染者比例、无症状感染者的假阴性率、无症状感染者的假阳性率、感染者数量、治愈人数、病死人数对防疫工作及疾病流行的直接途径系数分别为-0.968、-0.866、-0.957、-0.876、-0.843、0.765、0.784, 其中潜伏期的长度对防疫工作及疾病流行的直接影响最大。综合考虑各因素对防疫工作及疾病流行影响的总效应, 由简单相关系数可知其绝对值的大小排序可以表征其对防疫工作及疾病流行影响重要程度, 分析结果如表 6-14 所示 (注: 表中按影响程度从大到小排序, 排序越靠前者影响程度越大)。

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

表 6-14 影响防疫工作及疾病流行的主要因素的排序

排序	影响因素	影响方向
1	潜伏期的长度 LEN	-
2	无症状感染者的假阴性率 FNR	-
3	无症状感染者的假阳性率 FPR	-
4	无症状感染者的比例 PER	-
5	感染者数量 NUM	-
6	病死人数 DIE	+
7	治愈人数 REC	+

## 6.2.4 基于去除 LEN 的防疫工作及疾病流行综合预测模型

### ● 简单相关系数处理

考虑到前三种最重要因素的准确性，我们需要量化最重要因素对防疫工作和疾病流行产生的影响。处理的方法是，以得到的各个主要因素与防疫工作及疾病流行评估指数之间的简单相关系数为基础，将去除某一因素后简单相关系数占比与去除该因素前简单相关系数占比作差，以此来量化未知事件对于各个主要因素的影响。去除 LEN 前后占比之差表示如式 6-12。

$$sub_i = \frac{p_i}{\sum_{i=1}^5 |p_i|} - \frac{p_j}{\sum_{j=1}^6 |p_j|}, \quad i=1,2,\dots,6; j=1,2,\dots,7 \quad (\text{式 6-12})$$

式中  $p_i$  代表去除 LEN 后 PER、FNR、FPR、NUM、REC、DIE 的简单相关系数， $p_j$  代表去除 LEN 前 LEN、PER、FNR、FPR、NUM、REC、DIE 的简单相关系数，前后占比之差为  $sub_i$ 。根据上述公式处理结果如表 6-15 所示。

表 6-15 去除潜伏期长度(LEN)前后简单相关系数占比情况

自变量	去除 LEN 前各因素占比	去除 LEN 后各因素占比	前后各占比之差
LEN	0.159762337	0	0.159762337
PER	0.142927876	0.170104105	0.027176229
FNR	0.157946856	0.187978786	0.03003193
FPR	0.144578313	0.172068356	0.027490043
NUM	0.13913187	0.165586329	0.026454459
REC	0.126258458	0.150265174	0.024006715
DIE	0.129394289	0.15399725	0.024602961

### ● 基于修正后的简单相关系数对数据的处理

以前后占比之差  $sub_i$  为  $p_i$  ( $i$  代表 PER、FNR、FPR、NUM、REC、DIE) 系数，对原数据进行处理，部分数据如下表 6-16，全部数据见附录。

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

表 6-16 基于修正后简单相关系数的部分数据处理结果

PER	FNR	FPR	NUM	REC	DIE
0.004422021	0.000902709	0.000289523	0.370362424	0	0
0.002166599	0.000924521	0.000292151	0.581998095	0	0
0.002262169	0.000934981	0.000293966	0.952360519	0.024006715	0
0.003750053	0.000936092	0.000326461	1.084632814	0.048013431	0
0.004590187	0.000939109	0.000328292	1.798903203	0.048013431	0
0.003164738	0.000943117	0.000336437	2.11635671	0.048013431	0.024602961
0.004477799	0.000965544	0.000360182	2.407355757	0.096026861	0.024602961

## ● 预测结果及对比分析

通过上述处理，我们实现了 LEN 的准确性对主要因素影响的量化，结合问题一建立的基于遗传算法改进 BP 神经网络的防疫工作及疾病流行评估指数预测模型，对防疫工作及疾病流行评估指数进行预测。模型运行效果如下。

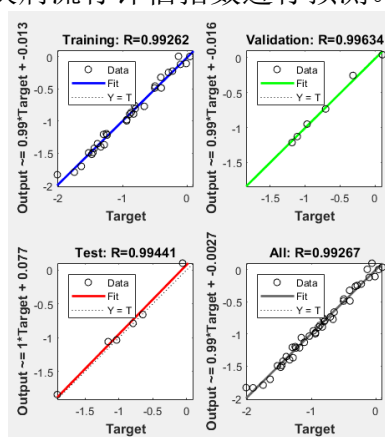


图 6-9 模型回归效果

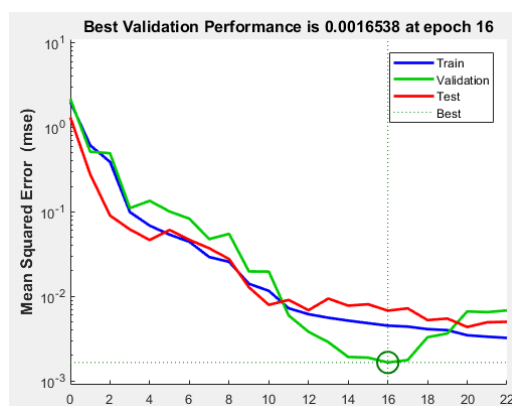


图 6-10 模型准确性检验

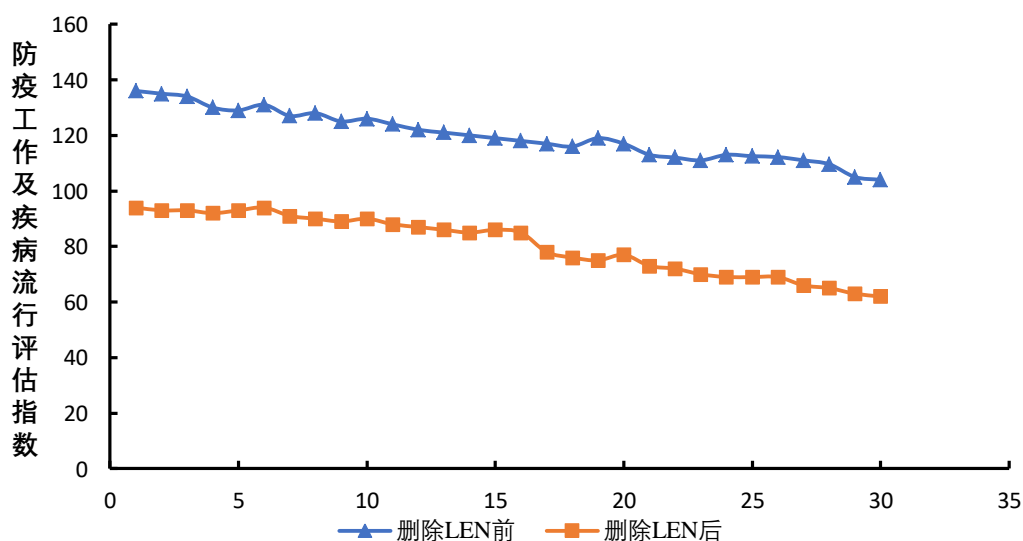


图 6-11 IM 去除前后预测结果对比

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

根据图 6-11 我们可以看出，去除潜伏期的长度 LEN 之后，防疫工作及疾病流行评估指数明显减低，说明若潜伏期的长度 LEN 数据不准确，大大降低潜伏时期发现感染者的概率，极有可能会使得防疫工作开展困难加大。

## 6.2.5 基于去除 FNR 的防疫工作及疾病流行综合预测模型

与上述基于去除 LEN 的煤炭价格预测模型原理相同，经过计算，可得预测结果如下表 6-17 所示：

表 6-17 去除无症状感染者的假阴性率(FNR)前后简单相关系数占比情况

自变量	去除 FNR 前各因素占比	去除 FNR 后各因素占比	前后各占比之差
LEN	0.159762337	0.189729518	0.029967181
PER	0.142927876	0.169737358	0.026809482
FNR	0.157946856	0	0.157946856
FPR	0.144578313	0.171697374	0.02711906
NUM	0.13913187	0.165229322	0.026097452
REC	0.126258458	0.1499412	0.023682741
DIE	0.129394289	0.153665229	0.02427094

模型运行效果如下图 6-12 所示：

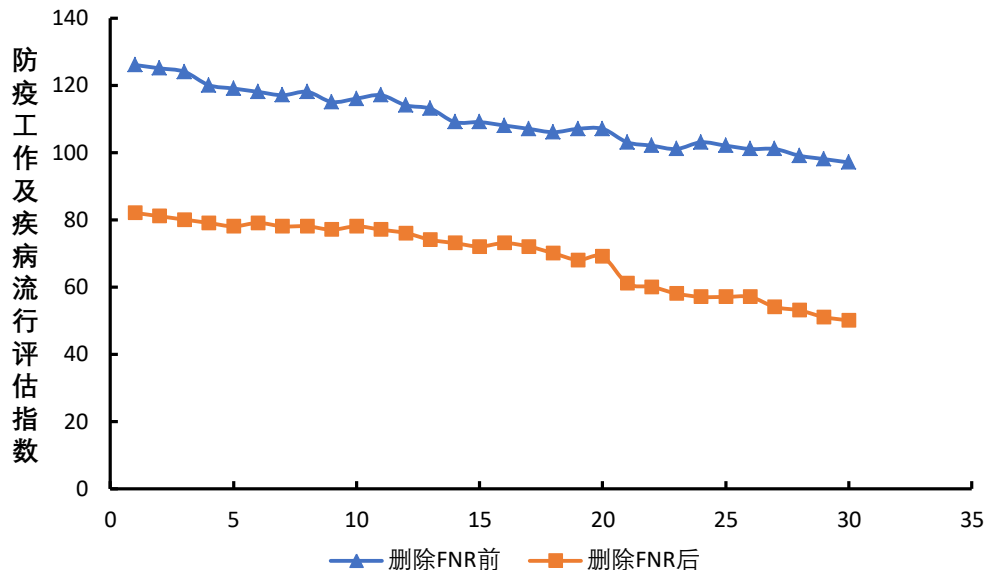


图 6-12 FNR 去除前后预测结果对比

根据图 6-12 我们可以看出，去除无症状感染者的假阴性率 FNR 因素之后，防疫工作及疾病流行评估指数呈现下降趋势。由于某些技术等方面的原因，无症状感染者的假阴性率 FNR 数据明显偏小，说明部分感染的群众未被检测出来，这会给社会带来极大的隐患，从而导致防疫工作开始更加困难。



# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## 6.2.6 基于去除 FPR 的防疫工作及疾病流行综合预测模型

与上述基于去除 LEN 的煤炭价格预测模型原理相同，经过计算，可得预测结果如下表 6-18 所示：

表 6-18 去除无症状感染者的假阳性率(FPR)前后简单相关系数占比情况

自变量	去除 FPR 前各因素占比	去除 FPR 后各因素占比	前后各占比之差
LEN	0.159762337	0.186764422	0.027002085
PER	0.142927876	0.1670847	0.024156824
FNR	0.157946856	0.184642099	0.026695243
FPR	0.144578313	0	0.144578313
NUM	0.13913187	0.162647116	0.023515246
REC	0.126258458	0.147597916	0.021339458
DIE	0.129394289	0.151263747	0.021869457

模型运行效果如下图 6-13 所示：

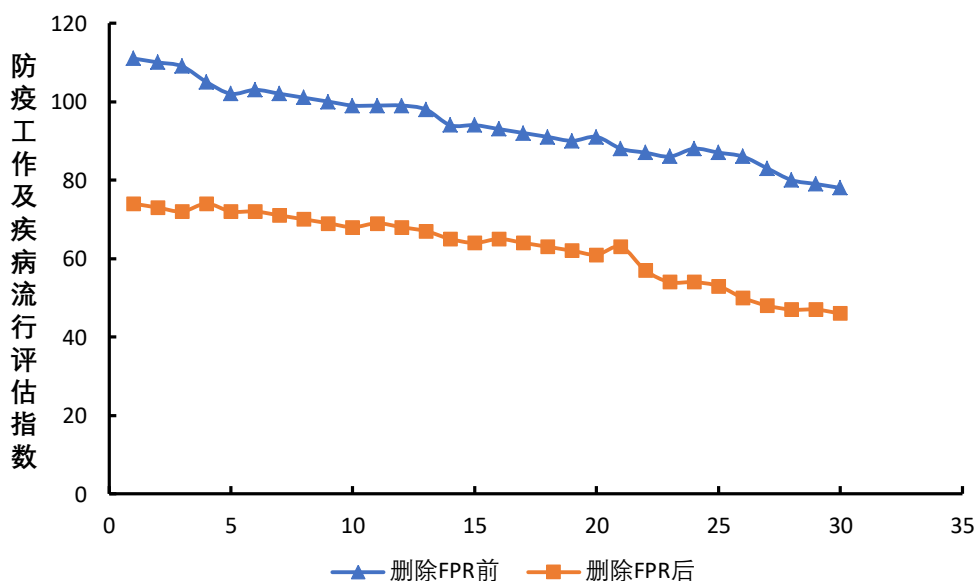


图 6-13 FPR 去除前后预测结果对比

根据图 6-13 可以看出，无症状感染者的假阳性率 FPR 删除前后的趋势和无症状感染者的假阴性率 FNR 基本相同，同样，无症状感染者的假阳性率 FPR 的降低也会增大防疫工作开展的难度。

## 6.3 问题三模型的建立与求解

体育赛事对体育产业发展具有重要的引领作用，因此重启体育赛事显得尤为重要。我们结合第一阶段建立的引入无症状感染者的 SEIR 模型，以北京市为例，结合北京市的 COVID-19 确诊病例和无症状感染者病例的数据，利用幂函数进行数据拟合预测各群体的人数，评估重启大型体育赛事的可能性，并分阶段给出北京市重启大型体育赛事的

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

时间表。

## 6.3.1 引入无症状感染者的 SEIR 模型

### ● 传统的 SEIR 模型

在传统的 SEIR 模型中，研究对象被分成 S、E、I、R 等 4 种类型。

(1) **S(Susceptible)**: 易感者，从未感染病毒的健康人群，但会因为接触感染者而感染病毒。

(2) **E(Exposed)**: 潜伏者，感染了病毒的易感者在出现症状之前所处的状态。

(3) **I(Infectives)**: 感染者，在 14 天潜伏期出现自我感知或可临床识别的症状与特征的病毒携带者，具有较强的传播能力，可以将病毒传播到接触的易感者。

(4) **R(Recovered)**: 移出者，感染者治疗完成痊愈或者死亡的人群。恢复者治愈之后具有抗体，即使与感染者接触也不会被传染。

传统 SEIR 模型示意图如图 6-14 所示。

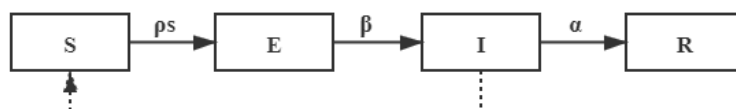


图 6-14 传统 SEIR 模型示意图

图中， $\rho$ 为一名感染者每一时刻平均接触到的人数； $s$ 为易感者所占所有人群的比例； $\beta$ 为每一时刻潜伏者中将会有占 $\beta$ 的人群转化为感染者，其中 $\beta$ 为平均潜伏期的倒数； $\alpha$ 表示治愈率和死亡率的总和。

传统 SEIR 模型存在着许多改进的空间，为了使其更加贴合 COVID-19 的传播规律。就如来说，通过考虑时变传播率，我们可以引入无症状感染者等更多类型的人群。

### ● 引入无症状感染者的 SEIR 改进模型

#### Step1. 人群类型修改

##### (1) E 重定义

被监测出有可疑迹象，实际上可能携带病毒也可能不携带病毒的人群。

##### (2) 引入 U(Unsymptomatic): 无症状感染者

一是：感染者核酸检测呈阳性，经过 14 天潜伏期的观察，均无任何可自我感知或可临床识别的症状与体征，始终为无症状感染状态。二是：感染者核酸检测呈阳性，采样时无任何可自我感知或可临床识别的症状与体征，但随后出现某种临床表现，即处于潜伏期的“无症状感染”状态。

U 与 E 在本模型中最大的区别在于，U 容易被认为是“健康的”，从而没有被采取任何强制措施，这种情况下他们的传播能力将远大于 E。

##### (3) I 重定义

依据国家“自行在家隔离”防控策略，除医务人员之外，其他人被感染几率极低。且因防护设备密不透风，感染可能性相对较低。由此，对 I 传播能力进行重定义，认为 I 不具有较高的传播能力。

#### Step2. 修正后 SEIR 模型的假设

(1) 依据重定义后的 I，假设 I 对 S 没有传播能力。

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

(2) 根据目前已收集到的数据, 全国易感者比例始终接近于 1 (非易感者人数在 30 万人以内), 对传播率的影响较小, 由此假设 U 和 E 的传播率是恒定的。

(3) 随着针对 COVID-19 治疗技术的研发和改善, 治愈率与死亡率必然发生变化, 由此假设治愈率和死亡率与时间相关。

(4) 假定 S 可以被 U 和 I 传染。

(5) E 中未携带病毒的人在 14 天 (医学观察期为 14 天) 后将变为 S, 因此假设每日  $1/14$  的 E 回到易感人群。

(6) 被治愈者再感染 COVID-19 的可能性极低, 由此假设被治愈者不会被传播。

## Step3. 修正后 SEIR 模型的建立

修正后的 SEIR 模型示意图如图 6-15 所示。

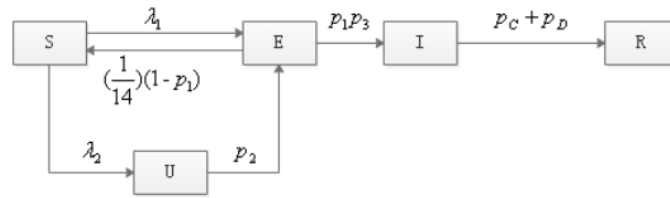


图 6-15 修正后的 SEIR 模型示意图

一位 R 在单位时间内可将  $\lambda_1$  个 S 变为 E, 而一名 U 在单位时间内传染  $\lambda_2$  个 S, 携带病毒的 E 在所有 E 中所占比例为  $p_1$ , 由此可得 S 变化方程为:

$$\frac{dS_t}{dt} = -\lambda_1 E_t - \lambda_2 U_t + \frac{1}{14}(1-p_1)E_t \quad (\text{式 6-13})$$

U 在易感人群中传播病毒后, 被传播病毒的 S 变为 U, 这些 U 如果在体温轨迹等监测中出现疑似点, 则有一定概率  $p_2$  被发现, 变为 E, 由此可得 U 变化方程为:

$$\frac{dU_t}{dt} = \lambda_2 U_t - p_2 U_t \quad (\text{式 6-14})$$

带病毒 E (比例为  $p_1$ ) 在单位时间内变为 I 的概率为  $p_3$ , 未带病毒 E (比例为  $1-p_1$ ) 在 14 天观察期后, 返回 I 中, 由此可得 E 变化方程为:

$$\frac{dE_t}{dt} = p_2 U_t + \lambda_1 E_t - p_1 p_3 E_t - \frac{1}{14}(1-p_1)E_t \quad (\text{式 6-15})$$

E 被确诊后变为 I, 而 I 因至于或死亡而变为 R, 被治愈者不会被传播, t 时刻的治愈率和死亡率分别为  $p_{c,t}$  和  $p_{d,t}$ , 由此可得 I 和 R 变化方程为:

$$\frac{dI_t}{dt} = p_1 p_3 E_t - (p_{c,t} + p_{d,t}) I_t \quad (\text{式 6-16})$$

$$\frac{dR_t}{dt} = (p_{c,t} + p_{d,t}) I_t \quad (\text{式 6-17})$$

## 6.3.2 北京市感染者预测结果

### Step1. 数据选取与分析

# 第十三届数学中国数学建模网络挑战赛

参赛队号 #32491

研究数据包括来源于北京市卫生健康委员会通报的确诊病例个案数量及详情、每日病例诊疗情况，北京市 COVID-19 确诊病例累计情况如图 6-16 所示：

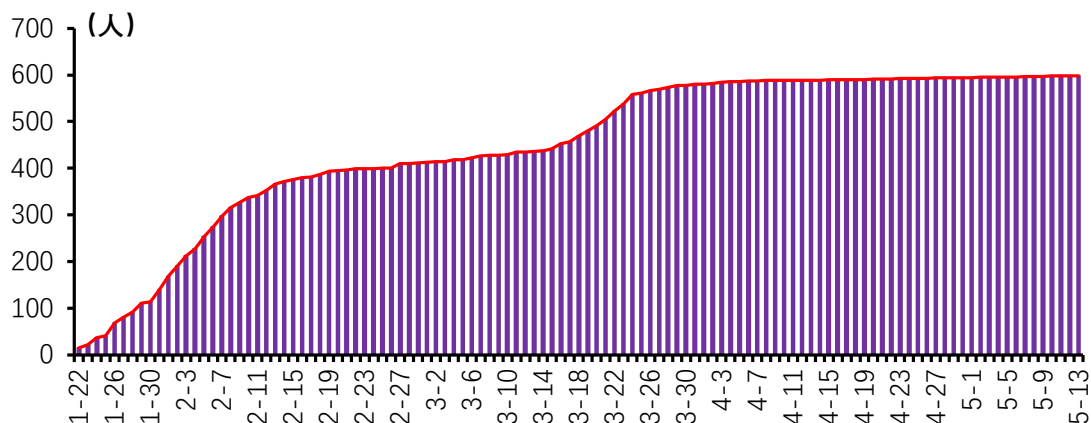


图 6-16 北京市 COVID-19 确诊病例累计情况

通过分析上图北京市 COVID-19 确诊病例累计情况，我们可以看出北京市的感染数量基本稳定，新增病例很少甚至为零。

根据北京市治愈病例和死亡病例的数据，可以得到如图 6-17、6-18 所示的北京市的治愈率和死亡率的时间演变图。

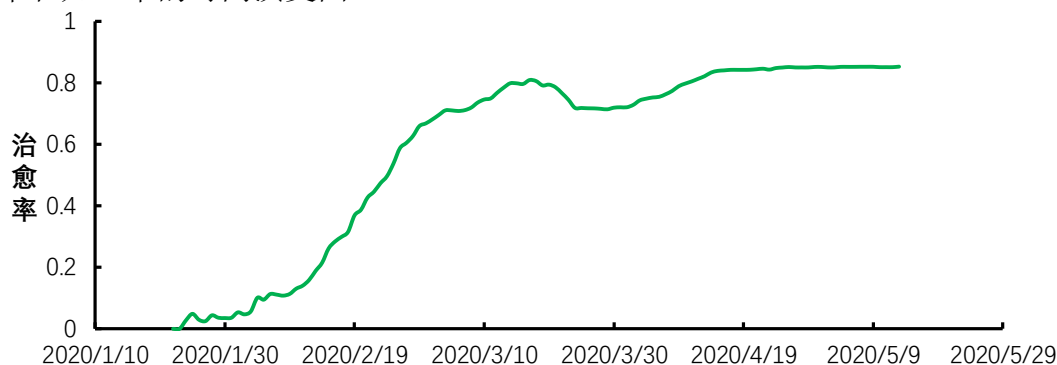


图 6-17 治愈率时间演变图

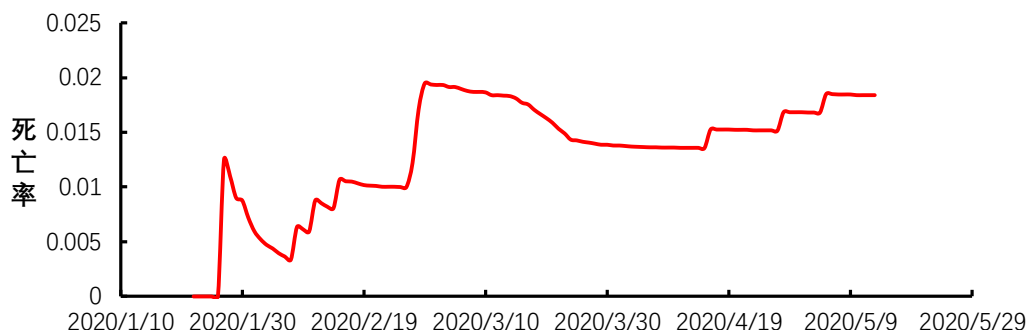


图 6-18 死亡率时间演变图

由上图可以看出，北京市治愈病例一直呈上升趋势，最后趋于平缓，说明感染者的数量也在降低；死亡率在 3 月一直持续走低，之后开始有些许上升，原因在于无症状感染者出现，导致传染的可能性进一步增大，但北京市死亡率整体较低。

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

## Step2. 数据拟合与分析

我们利用幂函数对北京市的治愈率和死亡率进行了拟合，并去除波动较大值得到了具体的拟合结果。 $p_{C,t}$ 和 $p_{D,t}$ 分别为当天新增出院人数和新增死亡人数除以上一天感染者的比率，结果如图 6-19、图 6-20 所示。

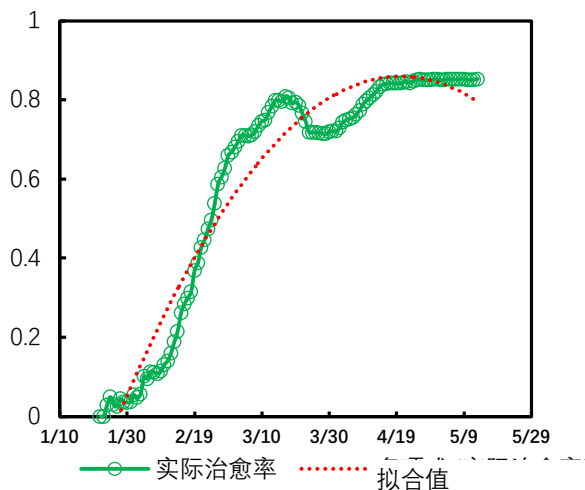


图 6-19 实际治愈率与拟合效果图

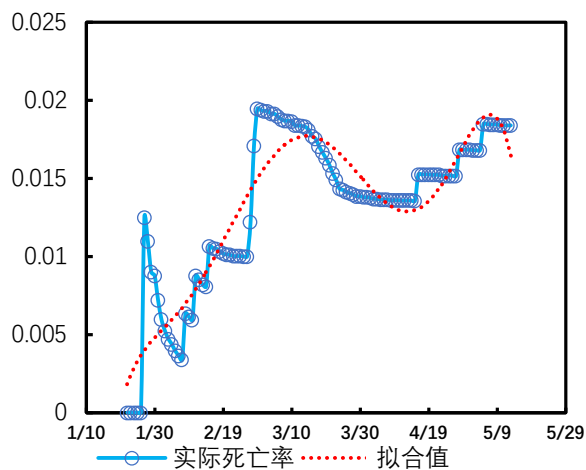


图 6-20 实际死亡率和拟合效果图

$p_{C,t}$ 和 $p_{D,t}$ 随时间的推移分别表现出明显的增长和下降趋势，拟合方程如表 6-19 所示。

表 6-19 拟合方程及系数

拟合方程	可决系数 $R^2$	调整可决系数
$p_{C,t} = 7.935 \times 10^{-4}t^{1.0380}$	0.9012	0.8996
$p_{D,t} = 1.137 \times 10^{-2}t^{-0.6234}$	0.9822	0.9817

由于所收集的数据以天为单位，因此考虑离散变化的情况，相关的微分方程转化成如下的差分方程：

$$S_{t+1} - S_t = -\lambda_1 E_t - \lambda_2 U_t + \frac{1}{14}(1 - p_1)E_t$$

$$U_{t+1} - U_t = \lambda_2 U_t - p_2 U_t$$

$$E_{t+1} - E_t = p_2 U_t + \lambda_1 E_t - p_1 p_3 E_t - \frac{1}{14}(1 - P_1)E_t$$

$$I_{t+1} - I_t = p_1 p_3 E_t - (p_{C,t} + p_{D,t})I_t$$

$$R_{t+1} - R_t = (p_{C,t} + p_{D,t})I_t$$

在进行疫情数据的拟合中，我们可以看到引入无症状感染者的 SEIR 模型相对于传统 SEIR 模型的优势更为明显，表 6-20 为拟合结果，由表 6-21 我们可以看出改进的 SEIR 模型拟合误差在 RMSE 和 MAPE 上分别降低了 70.75% 和 70.60%。

# 第十三届数学中国数学建模网络挑战赛

参赛队号 #32491

表 6-20 拟合结果

无症状给感染者的引入	$\lambda_1$	$\lambda_2$	$p_1$	$p_2$	$p_3$	$U_0$
引入	0.0239	0.7954	0.1321	0.9656	0.1232	69921
未引入	0.2137	—	0.4132	—	0.3382	—

表 6-21 拟合效果评价

无症状给感染者的引入	均方根为误差(RMSE)	平均相对误差(MAPE)/%
引入	1192.43	5.2452
未引入	4076.82	17.8413

## Step3. 模型的仿真预测

基于拟合好的 SEIR 模型，对北京市各群体的人数进行预测，如图 6-21 所示。

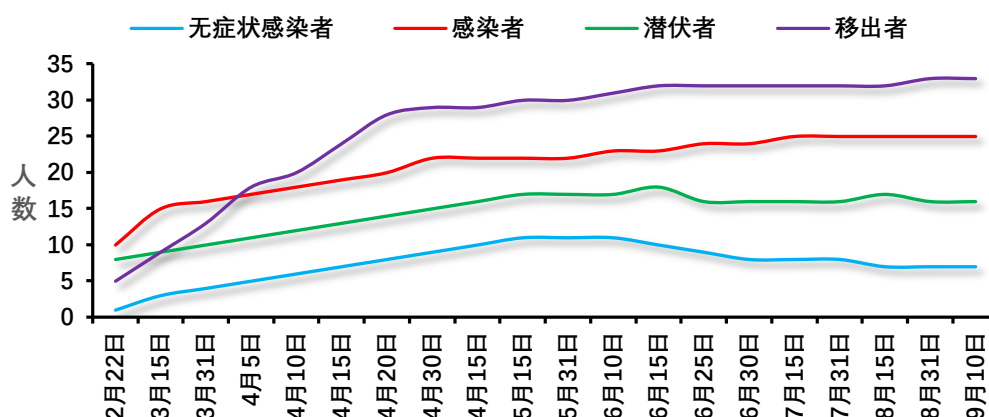


图 6-21 仿真预测结果

## 6.3.3 针对北京市的赛事重启安排

由图 6-21 我们可以得出，无症状感染者的数量在不断降低，整体感染者的数量在缓慢上升，但增幅不大。对于北京市开展大型体育赛事而言，风险仍然很高，因此如果重启赛事，需要做好防控措施。赛事重启时间表如表 6-22 所示。

表 6-22 赛事重启时间表

时间	赛事重启阶段
5 月末	无观众赛事
6 月中旬	采取限流的方式组织部分观众赛事
7、8 月左右	全部观众都可观看，但仍要佩戴好口罩，做好防护
9、10 月份之后	全面开放赛事



# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

## 6.4 疫情防控备忘录

尊敬的国际奥委会：

您好！

2020 年新型冠状病毒引发的肺炎疫情爆发席卷全球，在严重损害世界人民身体健康的同时，国内外体育赛事受到巨大冲击，随着东京奥运会的改期，疫情给全球体育赛事所带来的影响越来越严重和复杂。本团队本着人类命运共同体的原则，积极对各国出现二次高峰的风险性、参数准确性对防疫工作和疾病流行过程的影响、大型体育赛事的分阶段重启等问题开展研究，给出如下防控建议和降低风险的思路：

首先，当重启某项大型体育赛事时，我们需要评估赛事举办地区出现二次高峰的风险性以选择合适的时间与防疫策略。本团队在综合考虑了人口数、感染数量、病死人数、疫情持续时间、经济状况、医疗条件、人口密度、防疫政策等因素的情况下构建流行病流行程度评估指数  $F$ ，基于 GA-BP 神经网络模型对赛事所在地的疫情进行预测得到其第二次疫情高峰的预测时间，基于峰值的均差与时差从而得到风险指数  $R$ ，风险指数越大则举办地区出现二次高峰的风险性越高。

其次，为了减小因参数准确性而导致对体育赛事举办地区防疫工作和疾病流行过程的影响，以制定不同等级的并且合适的防疫策略。本团队基于因子分析构建防疫工作及疾病流行评估指数  $F_a$ ，通过通径分析对  $F_a$  影响显著的因素进行排序，将剔除最重要因素后通过 GA-BP 神经网络模型预测出的  $F_a$  与未剔除任何最重要因素时的  $F_a$  进行对比，从而可以得出因参数准确性而导致对防疫工作和疾病流行过程的影响，凭借不同的影响程度以指定合适的防疫策略。

再次，为保障参赛观众的安全性，保证赛事的正常进行。本团队引入无症状感染者的 SEIR 模型，实现对无症状感染者等各群体人数的预测，通过无症状感染者人数走势来规划重启时间表。

最后，我们团队认为，应当科学应对，化“危”为“机”。应进一步提升体育赛事管理服务水平，充分发挥体育社会组织的功能和作用，科学部署重大赛事的筹办工作，引导办赛主体全力开展自救措施，加强疫后体育赛事长期跟踪，及时调整策略，推进体育赛事运营和管理服务水平的提升，助力体育产业高质量发展。

2020 年 5 月 16 日





# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

## 七. 模型的评价

### ● 基于因子分析法的流行病流行程度评估指数模型

(1) **优点:** 因子分析法通过研究变量间的相关系数矩阵, 把这些变量间错综复杂的关系归结成少数几个综合因子, 由于归结出的因子个数少于原始变量的个数, 但是它们又包含原始变量的信息, 基于因子分析法构建流行病流行程度评估指数, 设计思路简单, 易于实现, 能够帮助地区快速对该地区的流行病流行程度进行定量判别以帮助该地区制定相应的防控措施提供参考。

(2) **缺点:** 对数据的正确性依赖程度较高, 在应用该方法前需要通过共线性检验, 对于样本量和指标数之间的关系需要提前把控。

### ● GA-BP 神经网络模型

(1) **优点:** 该模型预测精度更高且收敛速度更快; 预测均方误差得到了很大改善。

(2) **缺点:** 基于遗传算法优化的神经网络模型的局部处理不是很好, 且内部关系是复杂的非线性关系

### ● 通径分析

(1) **优点:** 可用于分析多个自变量与因变量之间的线性关系, 是回归分析的拓展, 可以处理较为复杂的变量关系。

(2) **缺点:** 其限制性较强对自变量与因变量之间的关系要求较高, 一般需要配合逐步回归、多元回归等统计学方法使用较为复杂。

## 八. 模型的应用与推广

### ● 基于因子分析法的流行病流行程度评估指数模型的应用与推广

1、可用来构建广告媒体传播能力的评价体系, 能够对经济状况、人口状况等因素进行综合考虑。

2、可用来构建一些需要对各种定量与定性相结合的综合因素综合考虑的评价体系。

### ● GA-BP 神经网络模型的应用与推广

用于更高精度的预测, 如天气预报、鱼群的迁徙预测等, 与实际情况进行比对, 修改训练参数, 进一步减少误差; 遗传算法在很多领域都十分有用, 可用于电力系统进行无功优化求解, 在水利土木工程发展中也能够进一步得到应用和推广。

### ● 通径分析的应用与推广

通径分析是一种将流行模型与多元回归相结合的因果关系综合分析方法。可应用于多基因遗传病学、社会学、经济学、农业及教育学等因素之间具有复杂关系的实际问题, 可以揭示要素之间互相影响的程度, 以及要素对所研究问题的直接与间接影响程度。

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

## 九. 参考文献

- [1]丁志华,冯猜猜,刘振华.我国煤炭价格波动及其因素研究——基于通径分析理论[J].价格理论与实践,2015(09):33-35.
- [2]龚正伟,刘星.新型冠状病毒肺炎疫情下基于人类命运共同体理念的我国体育治理方略[J].体育学研究,2020,34(02):34-40.
- [3]张华玲,王健,翟晓辉,向准,赵靖,刘俊峰.新型冠状病毒肺炎疫情下方舱医院党建工作思考[J].中国医院管理,2020,40(04):72-74.
- [4]钟秉枢,黄志剑,王凯,车冰清,宋昱.困境与应对: 聚焦新型冠状病毒肺炎疫情对体育事业的影响[J].体育学研究,2020,34(02):9-33+40.
- [5]黄海燕,刘蔚宇.新型冠状病毒肺炎疫情对体育赛事发展的影响研究[J].体育学研究,2020,34(02):51-58.
- [6]王雅伦,刘洋,吴珏.新型冠状病毒肺炎疫情对体育产业影响的预测研究[J].成都体育学院学报,2020,46(02):62-69.

## 附录

### 附录 1

问题二：基于修正后简单相关系数的部分数据处理结果（去除 LEN）

PER	FNR	FPR	NUM	REC	DIE
0.003996083	0.001002771	0.000382506	2.936444935	0.096026861	0.024602961
0.002759246	0.001007669	0.000396952	3.015808311	0.096026861	0.024602961
0.003389702	0.001067754	0.000408908	3.677169783	0.120033577	0.024602961
0.001812709	0.001083873	0.000413756	4.44434909	0.216060438	0.024602961
0.003295424	0.001094625	0.000419262	5.052801644	0.216060438	0.024602961
0.001695197	0.001100525	0.000459359	5.60834528	0.288080584	0.024602961
0.003627078	0.001105504	0.000471985	6.031616622	0.552154453	0.024602961
0.001788204	0.001107475	0.000479172	6.692978094	0.576161168	0.024602961
0.003209415	0.001110918	0.000496217	7.24852173	0.744208176	0.024602961
0.00346339	0.001124364	0.000499155	7.856974284	0.792221606	0.024602961
0.00251981	0.001188586	0.000514115	8.333154544	0.816228322	0.049205921
0.002739778	0.00120576	0.000525253	8.624153592	0.888248468	0.049205921
0.002705586	0.001246608	0.000527167	8.915152639	1.056295475	0.049205921
0.003260426	0.001259141	0.000537261	9.047424934	1.152322337	0.073808882
0.003239742	0.001262151	0.000550892	9.311969522	1.344376059	0.073808882
0.003788406	0.001266234	0.000566696	9.682331946	1.656463359	0.073808882
0.003368558	0.001278062	0.000576169	9.8410587	1.920537228	0.073808882
0.004322842	0.001297006	0.000579691	9.920422076	2.352658104	0.098411842
0.001784907	0.00129988	0.000588663	10.05269437	2.592725257	0.098411842
0.00407269	0.001304454	0.000598056	10.07914883	2.736765549	0.098411842
0.003913175	0.001333803	0.000639499	10.23787558	2.928819272	0.098411842

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

0.003721496	0.0013465	0.000645083	10.39660234	3.480973725	0.098411842
0.002463189	0.001348025	0.000662328	10.44951125	3.673027448	0.098411842
0.002110771	0.001355879	0.000672791	10.47596571	4.057134893	0.098411842
0.001390128	0.001369715	0.000686879	10.55532909	4.273195331	0.098411842
0.002779484	0.001373211	0.000695254	10.55532909	4.5372692	0.098411842
0.002473047	0.001386256	0.000703393	10.55532909	4.753329638	0.098411842
0.003262989	0.001399198	0.000703972	10.58178355	5.161443799	0.098411842
0.002576638	0.001426744	0.000730828	10.58178355	5.641578106	0.098411842
0.001560185	0.001427224	0.000741923	10.84632814	5.953665405	0.123014803
0.004164773	0.001429268	0.00076164	10.84632814	6.169725844	0.172220724
0.001709011	0.001429425	0.000765496	10.8727826	6.505819858	0.196823684
0.003287783	0.001436155	0.000771856	10.92569151	6.625853435	0.196823684
0.002367472	0.001471481	0.000788316	10.95214597	6.769893727	0.196823684
0.003711007	0.001506896	0.00078966	10.95214597	6.913934019	0.196823684
0.002765511	0.001513983	0.00079933	11.05796381	7.129994457	0.196823684
0.002896038	0.001520546	0.000810616	11.05796381	7.129994457	0.196823684
0.001441591	0.001550285	0.000834326	11.16378164	7.178007888	0.196823684
0.002054803	0.001559294	0.000838126	11.26959948	7.274034749	0.196823684
0.004373693	0.001570739	0.000841433	11.3225084	7.394068326	0.196823684
0.002325652	0.00157284	0.000841777	11.3225084	7.562115334	0.196823684
0.003072489	0.0015906	0.000842176	11.34896286	7.68214891	0.196823684
0.002833318	0.001593689	0.000850936	11.50768961	7.826189202	0.196823684
0.002939091	0.001598371	0.000863279	11.50768961	8.018242925	0.196823684
0.001676291	0.001627471	0.000889752	11.53414407	8.210296648	0.196823684
0.003325312	0.001672979	0.000891376	11.56059853	8.378343655	0.196823684
0.002958244	0.001712218	0.000896403	11.69287082	8.474370517	0.196823684
0.003935151	0.001731056	0.00090284	11.95741541	8.642417524	0.196823684
0.002623843	0.001733349	0.000903507	12.06323324	8.858477962	0.196823684
0.001847252	0.001738054	0.000910256	12.40714121	9.0745384	0.196823684
0.002891393	0.001746868	0.000910725	12.69814026	9.122551831	0.196823684
0.004279718	0.001762975	0.000922351	12.98913931	9.362618984	0.196823684
0.002925082	0.0017711	0.000927181	13.33304727	9.506659276	0.196823684
0.00185164	0.001809267	0.000970657	13.80922753	9.602686138	0.196823684
0.003054469	0.001809645	0.000978224	14.20604441	9.602686138	0.196823684
0.002646912	0.001839948	0.000991133	14.76158805	9.626692853	0.196823684
0.003255448	0.00184608	0.000998408	14.84095143	9.674706284	0.196823684
0.003535582	0.001847929	0.001038182	14.97322372	9.74672643	0.196823684
0.004033576	0.001850666	0.001050808	15.0525871	9.794739861	0.196823684
0.001842388	0.001852461	0.001050856	15.15840493	9.842753291	0.196823684
0.003797573	0.001873593	0.001060606	15.26422277	9.890766722	0.196823684

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

0.004595118	0.001883097	0.001066871	15.26422277	9.962786868	0.196823684
0.002540965	0.001885371	0.001068725	15.34358614	10.03480701	0.196823684
0.003438437	0.00191348	0.00106949	15.34358614	10.03480701	0.196823684
0.001890066	0.001918483	0.001086757	15.39649506	10.17884731	0.196823684
0.002463728	0.001938264	0.001090823	15.44940398	10.41891446	0.196823684
0.003052028	0.001944501	0.001092235	15.47585844	10.51494132	0.196823684
0.002143176	0.001946945	0.001114651	15.5023129	10.58696147	0.196823684
0.002842685	0.001998278	0.001114876	15.52876736	10.6349749	0.196823684
0.003489535	0.002009921	0.001125164	15.52876736	10.75500847	0.196823684
0.002610802	0.002017005	0.001146708	15.55522182	10.92305548	0.196823684
0.001724142	0.002021981	0.00116995	15.55522182	11.13911592	0.196823684
0.004598032	0.00205411	0.001200883	15.55522182	11.2591495	0.196823684
0.003558904	0.00206365	0.001208239	15.58167627	11.37918307	0.196823684
0.0014697	0.002121383	0.001221955	15.58167627	11.49921665	0.196823684
0.003457334	0.002138865	0.001225487	15.58167627	11.61925023	0.196823684
0.002174227	0.0021389	0.001229479	15.58167627	11.78729723	0.196823684
0.001585627	0.002152397	0.001239594	15.60813073	11.8833241	0.196823684
0.002409277	0.002157995	0.001242794	15.60813073	11.90733081	0.221426645
0.001946839	0.002165781	0.001251992	15.60813073	11.93133753	0.221426645
0.003075868	0.002194798	0.001286927	15.60813073	11.93133753	0.221426645
0.002374751	0.002208673	0.001294073	15.60813073	11.93133753	0.221426645
0.004487013	0.002210949	0.001295973	15.63458519	11.95534424	0.221426645
0.00432292	0.002213855	0.001300104	15.63458519	11.97935096	0.221426645
0.003252565	0.00221651	0.001318999	15.63458519	12.00335767	0.221426645
0.004117879	0.002239307	0.001319752	15.68749411	12.00335767	0.221426645
0.001945091	0.002276143	0.001333759	15.68749411	12.07537782	0.221426645
0.004087796	0.002276273	0.001343709	15.68749411	12.09938453	0.221426645
0.001700875	0.00227746	0.001354388	15.68749411	12.12339125	0.221426645
0.004056821	0.002293803	0.001359344	15.71394857	12.12339125	0.221426645
0.001847617	0.002330332	0.001361958	15.71394857	12.12339125	0.246029606
0.004008647	0.002349077	0.00138831	15.71394857	12.12339125	0.246029606
0.004036712	0.002366484	0.001395521	15.71394857	12.14739796	0.246029606
0.002360842	0.002368268	0.001423148	15.71394857	12.14739796	0.246029606
0.001502822	0.002374792	0.001477491	15.74040303	12.14739796	0.246029606
0.00178197	0.002395337	0.001511868	15.74040303	12.14739796	0.246029606
0.00365321	0.002421515	0.001547029	15.74040303	12.17140468	0.246029606
0.003608699	0.002423348	0.001550564	15.74040303	12.17140468	0.270632566
0.004441601	0.002425348	0.001579307	15.74040303	12.17140468	0.270632566
0.004230773	0.002433657	0.001589853	15.76685749	12.1954114	0.270632566
0.004388795	0.002440324	0.001596085	15.76685749	12.1954114	0.270632566

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

0.00202949	0.0024538	0.001619549	15.76685749	12.1954114	0.270632566
0.003066039	0.00249521	0.001627114	15.8197664	12.21941811	0.270632566
0.002503742	0.002495579	0.001634687	15.8197664	12.21941811	0.270632566
0.001748999	0.002532005	0.001634845	15.8197664	12.21941811	0.270632566
0.002832524	0.002549726	0.001643255	15.8197664	12.24342483	0.270632566

## 附录二 BP 神经网络预测模型代码

```
function [y1] = myNeuralNetworkFunction(x1)
%MYNEURALNETWORKFUNCTION neural network simulation function.
%
% Generated by Neural Network Toolbox function genFunction, 02-May-2020 11:18:51.
%
% [y1] = myNeuralNetworkFunction(x1) takes these arguments:
%   x = Qx6 matrix, input #1
% and returns:
%   y = Qx1 matrix, output #1
% where Q is the number of samples.

%#ok< *RPMTO>

% ===== NEURAL NETWORK CONSTANTS =====

% Input 1
x1_step1.xoffset =
[505365.5094;3.05577934;180738.6513;-1114.110091;44.93985933;503.5792857];
x1_step1.gain =
[6.72385096159665e-06;5.15334965576101;3.00413226444048e-06;0.000431324467496717;0.03366261
55006605;0.00625567729692304];
x1_step1.ymin = -1;

% Layer 1
b1 =
[2.4404118243219996565;-1.7360289282513752251;1.8916562356582657944;-0.6104625677309148956
7;0.17136431674705332462;-0.25953009407825583521;-1.1884029936657143356;1.4019781554879877
65;-1.648824811263809087;-2.3589576588880696306];
IW1_1 = [-0.60932359739893104944 0.2748780684330002777 -0.86152311261018044153
0.28305032474722147118 -0.59900184364083164379
-1.0299378578105424786;0.28944325836906864557 1.1348005499352376013
-0.28709681119087110401 -0.32722630647769412882 0.87410634743712123385
0.88916046724878761154;0.13192481407691333817 -0.17854811719505700451
0.03058125870947183203 -0.87329200094873182803 0.60147320950524296723
-1.6082014292580646408;0.55179263952333956755 -0.21430956965318642693
-0.58882635387503312607 1.568615352987206224 -0.72837118504828546328
-0.47194373987991278252;-0.9379234079780091049 0.94765284782733394398
-0.85407290076122577283 -0.98613475615979673261 -0.083498687015167438741
-0.41504728178223526669;-0.55406245044003532563 0.65676703433985894609
-0.59474251711439984902 0.73648462398473601986 1.2508347885715842551
0.8438562762993864208;-0.28372969749528215155 -0.38247166530556625519
0.77196776848134296412 -0.89943746075181851829 0.82326600107546576979
-1.2689086268813671854;0.68197622339071806152 0.58563692953155088716
0.98341126262214284548 -0.75543590094834045523 -0.39990872778261427056
```

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

```
1.1022653170184542049;-0.20777074508787823914 -0.99848830429831370026
-0.4135133707809630077 0.078937589534708807459 -0.87976951813363502009
-0.84963575062426699436;-0.35921606325369320745 -0.84067000832498350871
0.1419462184669142979 0.50319270368128121707 -0.64265181232956403168
-1.1044588132876800302];
```

```
% Layer 2
b2 = -0.16475134741574745667;
LW2_1 = [0.056244269641068489474 -0.39322792349192137173 -0.42794158700329898437
0.01051227322826289029 0.07472190217336566187 -0.19372907040630482634
-0.56454007071687473118 0.16150780983205356711 0.58113007790492154925
-0.012549875593345530961];
```

```
% Output 1
y1_step1.ymin = -1;
y1_step1.gain = 0.0136035435293749;
y1_step1.xoffset = 475;
```

```
% ===== SIMULATION =====
```

```
% Dimensions
Q = size(x1,1); % samples
```

```
% Input 1
x1 = x1';
xp1 = mapminmax_apply(x1,x1_step1);
```

```
% Layer 1
a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*xp1);
```

```
% Layer 2
a2 = repmat(b2,1,Q) + LW2_1*a1;
```

```
% Output 1
y1 = mapminmax_reverse(a2,y1_step1);
y1 = y1';
end
```

```
% ===== MODULE FUNCTIONS =====
```

```
% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings)
y = bsxfun(@minus,x,settings.xoffset);
y = bsxfun(@times,y,settings.gain);
y = bsxfun(@plus,y,settings.ymin);
end
```

```
% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n,~)
a = 2 ./ (1 + exp(-2*n)) - 1;
end
```

```
% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings)
```

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

```
x = bsxfun(@minus,y,settings.ymin);
x = bsxfun(@rdivide,x,settings.gain);
x = bsxfun(@plus,x,settings.xoffset);
end

% =====fits=====
function createfigure(X1, YMatrix1, X2, Y1, X3, Y2, X4, Y3)
%CREATEFIGURE(X1, YMatrix1, X2, Y1, X3, Y2, X4, Y3)
% X1: x 数据的向量
% YMATRIX1: y 数据的矩阵
% X2: x 数据的向量
% Y1: y 数据的向量
% X3: x 数据的向量
% Y2: y 数据的向量
% X4: x 数据的向量
% Y3: y 数据的向量

% 创建 figure
figure1 = figure('Tag','TRAINING_PLOTPERFORM','NumberTitle','off',...
    'Name','Neural Network Training Performance (plotperform), Epoch 61, Validation stop.');
```

% 创建 axes

```
axes1 = axes('Parent',figure1);
hold(axes1,'on');
```

% 使用 semilogy 的矩阵输入创建多行

```
semilogy1 = semilogy(X1,YMatrix1,'LineWidth',2,'Parent',axes1);
set(semilogy1(1),'DisplayName','Train','Color',[0 0 1]);
set(semilogy1(2),'DisplayName','Validation','Color',[0 0.8 0]);
set(semilogy1(3),'DisplayName','Test','Color',[1 0 0]);
```

% 创建 semilogy

```
semilogy(X2,Y1,'DisplayName','Best','LineStyle',':','Color',[0 0.48 0]);
```

% 创建 semilogy

```
semilogy(X3,Y2,'MarkerSize',16,'Marker','o','LineWidth',1.5,...
    'LineStyle','none',...
    'Color',[0 0.48 0]);
```

% 创建 semilogy

```
semilogy(X4,Y3,'LineStyle',':','Color',[0 0 0]);
```

% 创建 ylabel

```
ylabel('Mean Squared Error (mse)','FontWeight','bold','FontSize',12);
```

% 创建 xlabel

```
xlabel('61 Epochs','FontWeight','bold','FontSize',12);
```

% 创建 title

```
title('Best Validation Performance is 86.7188 at epoch 55',...
    'FontWeight','bold',...
```



# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

```
'FontSize',12);

% 取消以下行的注释以保留坐标区的 X 范围
% xlim(axes1,[0 61]);
% 取消以下行的注释以保留坐标区的 Y 范围
% ylim(axes1,[0.9 110000]);
% 设置其余坐标区属性
set(axes1,'YMinorTick','on','YScale','log');
% 创建 legend
legend(axes1,'show');

% =====拟合=====
function createfigure(X1, YMatrix1, X2, Y1, X3, YMatrix2, X4, X5, YMatrix3, X6, YMatrix4, X7)
% CREATEFIGURE(X1, YMatrix1, X2, Y1, X3, YMatrix2, X4, X5, YMatrix3, X6, YMatrix4, X7)
% X1: x 数据的向量
% YMATRIX1: y 数据的矩阵
% X2: x 数据的向量
% Y1: y 数据的向量
% X3: x 数据的向量
% YMATRIX2: y 数据的矩阵
% X4: x 数据的向量
% X5: x 数据的向量
% YMATRIX3: y 数据的矩阵
% X6: x 数据的向量
% YMATRIX4: y 数据的矩阵
% X7: x 数据的向量

% 创建 figure
figure1 = figure('Tag','TRAINING_PLOTREGRESSION','NumberTitle','off',...
    'Name','Neural Network Training Regression (plotregression), Epoch 61, Validation stop.');
```

```
% 创建 subplot
subplot1 = subplot(2,2,1,'Parent',figure1);
hold(subplot1,'on');

% 使用 plot 的矩阵输入创建多行
plot1 = plot(X1,YMatrix1,'Parent',subplot1);
set(plot1(1),'DisplayName','Y = T','LineStyle',':','Color',[0 0 0]);
set(plot1(2),'DisplayName','Fit','LineWidth',2,'Color',[0 0 1]);

% 创建 plot
plot(X2,Y1,'Parent',subplot1,'DisplayName','Data','Marker','o',...
    'LineStyle','none',...
    'Color',[0 0 0]);

% 创建 ylabel
ylabel('Output ~ 0.9*Target + 56','FontWeight','bold','FontSize',12);

% 创建 xlabel
xlabel('Target','FontWeight','bold','FontSize',12);
```

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

---

```
% 创建 title
title('Training: R=0.94858','FontWeight','bold','FontSize',12);

% 取消以下行的注释以保留坐标区的 X 范围
% xlim(subplot1,[476.15 631.166399304792]);
% 取消以下行的注释以保留坐标区的 Y 范围
% ylim(subplot1,[476.15 631.166399304792]);
box(subplot1,'on');
axis(subplot1,'square');
% 创建 legend
legend1 = legend(subplot1,'show');
set(legend1,'Location','northwest');

% 创建 subplot
subplot2 = subplot(2,2,2,'Parent',figure1);
hold(subplot2,'on');

% 使用 plot 的矩阵输入创建多行
plot2 = plot(X3,YMatrix2,'Parent',subplot2);
set(plot2(1),'DisplayName','Y = T','LineStyle','-', 'Color',[0 0 0]);
set(plot2(2),'DisplayName','Fit','LineWidth',2,'Color',[0 1 0]);

% 创建 plot
plot(X4,Y1,'Parent',subplot2,'DisplayName','Data','Marker','o',...
     'LineStyle','none',...
     'Color',[0 0 0]);

% 创建 ylabel
ylabel('Output ~ = 0.78*Target + 1.3e+02','FontWeight','bold','FontSize',12);

% 创建 xlabel
xlabel('Target','FontWeight','bold','FontSize',12);

% 创建 title
title('Validation: R=0.92434','FontWeight','bold','FontSize',12);

% 取消以下行的注释以保留坐标区的 X 范围
% xlim(subplot2,[475 631.166399304792]);
% 取消以下行的注释以保留坐标区的 Y 范围
% ylim(subplot2,[475 631.166399304792]);
box(subplot2,'on');
axis(subplot2,'square');
% 创建 legend
legend2 = legend(subplot2,'show');
set(legend2,'Location','northwest');

% 创建 subplot
subplot3 = subplot(2,2,3,'Parent',figure1);
hold(subplot3,'on');

% 使用 plot 的矩阵输入创建多行
plot3 = plot(X5,YMatrix3,'Parent',subplot3);
```

# 第十三届数学中国数学建模网络挑战赛

参赛队号 # 32491

---

```
set(plot3(1),'DisplayName','Y = T','LineStyle',':', 'Color',[0 0 0]);
set(plot3(2),'DisplayName','Fit','LineWidth',2,'Color',[1 0 0]);

% 创建 plot
plot(X6,Y1,'Parent',subplot3,'DisplayName','Data','Marker','o',...
     'LineStyle','none',...
     'Color',[0 0 0]);

% 创建 ylabel
ylabel('Output  $\approx 0.88 \cdot \text{Target} + 66$ ','FontWeight','bold','FontSize',12);

% 创建 xlabel
xlabel('Target','FontWeight','bold','FontSize',12);

% 创建 title
title('Test: R=0.92992','FontWeight','bold','FontSize',12);

% 取消以下行的注释以保留坐标区的 X 范围
% xlim(subplot3,[477.69 631.166399304792]);
% 取消以下行的注释以保留坐标区的 Y 范围
% ylim(subplot3,[477.69 631.166399304792]);
box(subplot3,'on');
axis(subplot3,'square');
% 创建 legend
legend3 = legend(subplot3,'show');
set(legend3,'Location','northwest');

% 创建 subplot
subplot4 = subplot(2,2,4,'Parent',figure1);
hold(subplot4,'on');

% 使用 plot 的矩阵输入创建多行
plot4 = plot(X3,YMatrix4,'Parent',subplot4);
set(plot4(1),'DisplayName','Y = T','LineStyle',':', 'Color',[0 0 0]);
set(plot4(2),'DisplayName','Fit','LineWidth',2,'Color',[0.4 0.4 0.4]);

% 创建 plot
plot(X7,Y1,'Parent',subplot4,'DisplayName','Data','Marker','o',...
     'LineStyle','none',...
     'Color',[0 0 0]);

% 创建 ylabel
ylabel('Output  $\approx 0.89 \cdot \text{Target} + 64$ ','FontWeight','bold','FontSize',12);

% 创建 xlabel
xlabel('Target','FontWeight','bold','FontSize',12);

% 创建 title
title('All: R=0.94342','FontWeight','bold','FontSize',12);

% 取消以下行的注释以保留坐标区的 X 范围
% xlim(subplot4,[475 631.166399304792]);
```

# 第十三届中国数学建模网络挑战赛

参赛队号 # 32491

---

```
% 取消以下行的注释以保留坐标区的 Y 范围
% ylim(subplot4,[475 631.166399304792]);
box(subplot4,'on');
axis(subplot4,'square');
% 创建 legend
legend4 = legend(subplot4,'show');
set(legend4,'Location','northwest');

% uicontrol 当前不支持代码生成，请输入正确输入语法对应的 'doc uicontrol'
% 可以使用 GUIDE 来为 uicontrol 生成代码。有关详细信息，请输入 'doc guide'
% uicontrol(...);
```