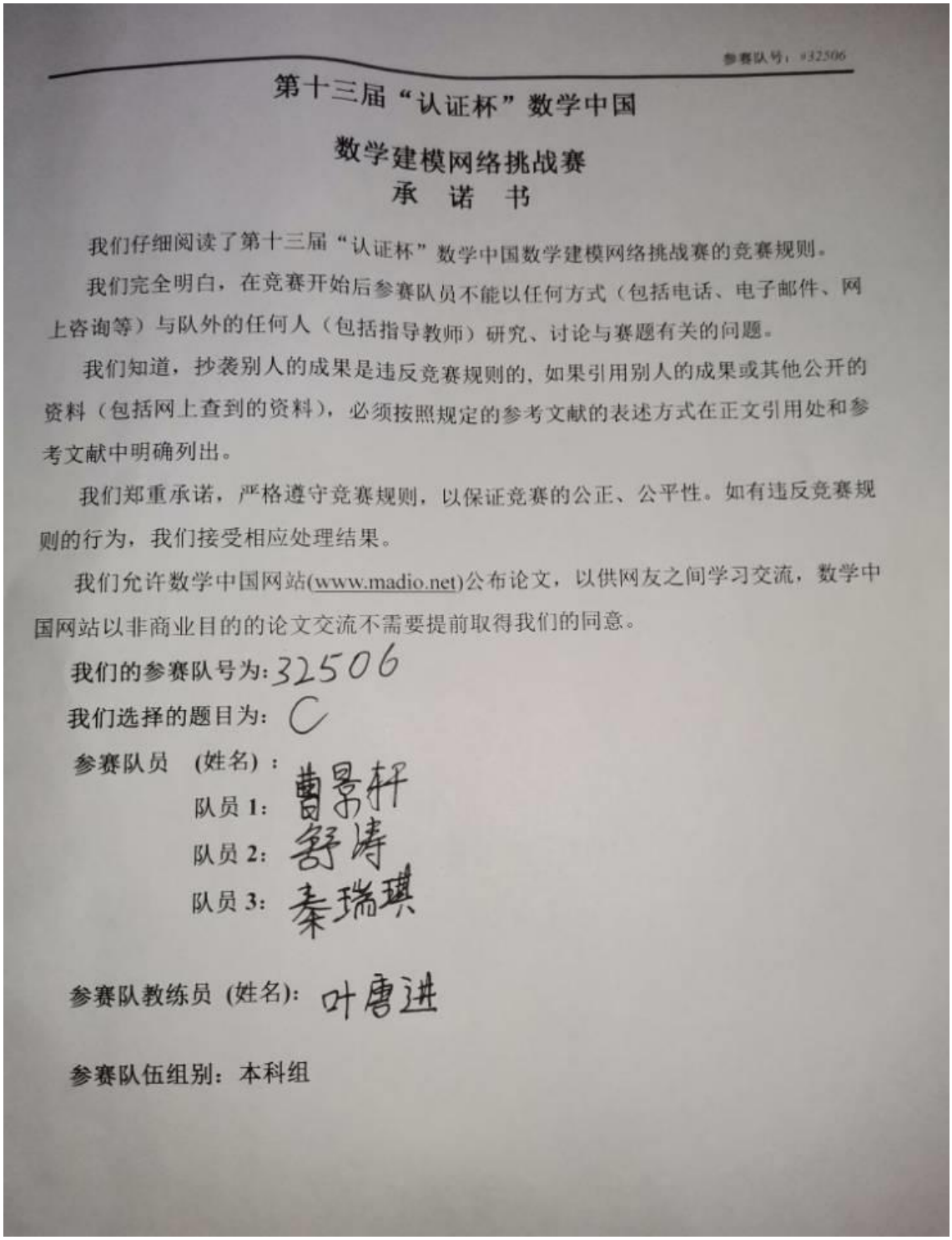

参赛队号：#32506

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)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：32506

我们选择的题目为：C

参赛队员 (姓名)：

队员 1：曹景轩

队员 2：舒涛

队员 3：秦瑞琪

参赛队教练员 (姓名)：叶唐进

参赛队伍组别：本科组

参赛队号：32506

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）：

32506

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
地址：数学中国数学建模网络挑战赛组委会　　网址：www.tzmcm.cn
电话：0471-4969085　　邮编：010021　　Email：service@tzmcm.cn

2020 年第十三届“认证杯”数学中国数学建模网络挑战赛第二阶段论文

题　　目　基于数学模型分析新型冠状病毒肺炎

关 键 词　NARX、随机森林、新型冠状病毒、疫情拐点预测、防疫建议

摘　　要：

2020年3月12日，世界卫生组织(WHO)宣布，席卷全球的冠状病毒引发的病毒性肺炎(COVID-19)是一种大流行病；建立新型冠状病毒(COVID-19)拐点预测模型为防治第二次大爆发，促进企业复工复产及促进防疫工作具有重要作用。

本文建立了基于 NARX 神经网络的传染病拐点预测模型，用于预测新型冠状病毒的未来发展趋势，促进企业复工复产。通过利用 2020 年 2 月 7 日-2 月 19 日中国新增确诊患者数据来验证模型的可信性。为了准确测量新型冠状病毒影响因子，本文建立就基于随机森林的影响指标提取模型，用于精准测量新型冠状病毒影响因子，促进防疫工作的开展。结合第一阶段模型及问题一的分析，充分考虑我国的疫情现状，分析了重启大型体育赛事的可能性，并设计了分阶段重启时间表。最后结合问题一、二、三的分析，为有关防疫部门设计了一份防疫防控备忘录。

针对问题一，本文利用现有的新型冠状病毒肺炎(COVID-19)数据，选取了中国、加拿大和澳大利亚三个国家，利用 NARX 神经网络预测了三个国家未来每日新增病例，来评估三国出现第二次高峰的风险程度，并给出了复工复产的政策性建议。

针对问题二：本文通过查阅相关文献，确定新型冠状病毒影响指标；通过利用随机森林算法，求取影响因子的特征值，并对其进行排序。讨论其中影响因子最高的指标，通过分析其作用，为防疫工作献计献策。

针对问题三：本文结合第一阶段的模型及前两个问题的分析，评估了我国重启大型体育赛事的可能性，为设计了分阶段重启的时间表。

针对问题四：结合前三问的分析，为重启一些大型体育赛事，本文设计了疫情防控备忘录。

参赛队号：32506

所选题目：C 题

参赛密码 _____

（由组委会填写）

Abstract

On 12 March 2020, the world health organization (WHO) declared that viral pneumonia caused by the coronavirus sweeping the world (covid-19) was a pandemic;The establishment of a novel coronavirus(covid-19) inflection point prediction model plays an important role in preventing and controlling the second outbreak, facilitating the resumption of work and production of enterprises and epidemic prevention.

A novel coronavirus inflection point prediction model based on NARX neural network was established to predict the future development trend of novel coronavirus and to promote the resumption of work and production.The credibility of the model was verified by using the data of newly diagnosed patients in China from February 7 to February 19, 2020.In order to accurately measure novel coronavirus factors, an extraction model based on random forest was established to accurately measure novel coronavirus factors and promote the development of epidemic prevention.Based on the first-stage model and the analysis of problem 1, the possibility of restarting large-scale sports events was analyzed, and the schedule of restarting large-scale sports events was designed.Finally, combining the analysis of question 1, 2 and 3, a memorandum of epidemic prevention and control was designed for the relevant epidemic prevention departments.

In response to question 1, this paper used the existing covid-19 data to select three countries -- China, Canada and Australia. NARX neural network was used to predict the daily new cases in the three countries in the future, so as to assess the risk level of the second peak in the three countries and give policy Suggestions for resuming work and production.

For question 2: the novel coronavirus influence index was determined by referring to relevant literature.By using the random forest algorithm, the eigenvalues of the influence factors are obtained and sorted.The index with the highest influence factor is discussed, and its function is analyzed to offer Suggestions for epidemic prevention.

Based on the model of the first stage and the analysis of the first two problems, this paper evaluates the possibility of restarting large-scale sports events in China, and designs a timetable for restarting the games in stages.

Aiming at question 4: based on the analysis of the first three questions, this paper designed a memorandum on epidemic prevention and control in order to restart some large-scale sports events.

目录

一、问题重述.....1
1.1 问题的背景与意义.....1
1.2 问题的提出.....2
二、模型假设.....2
三、符合说明.....3
四、问题分析.....3
五、问题一模型的建立与求解.....4
5.1 三个国家的疫情发展现状及其特点.....5
5.2 NARX 神经网络原理.....7
5.2.1 建立预测模型.....8
5.2.2 预测结果检验.....11
六、问题二分析与模型构建.....12
6.1 随机森林原理.....13
6.2 参数不准确的影响.....15
七、问题三体育赛事重启评估.....16
八、问题四的备忘录.....18
九、参考文献.....20

一、问题重述

1.1 问题的背景与意义

2020年3月12日，世界卫生组织(WHO)宣布，席卷全球的冠状病毒引发的病毒性肺炎(COVID-19)是一种大流行病。世界卫组织上一次宣布大流行是在2009年的H1N1流感爆发期间，该病感染了世界近四分之一的人口。但是，当时该决定因制造了不必要的恐慌而受到批评。SARS尽管影响了26个国家，但仍未被认为是大流行病，MERS也没有被认为是大流行病。世界卫组织表示，大流行是“新疾病的全球传播”。对于达到大流行水平与否，当下没有定量的严格标准，也没有触发该定义的病例或死亡数量阈值。也就是说“大流行”特征所指的不是疾病的严重性，而是疾病传播的广泛程度。目前，在全球已有超过200个国家/地区报告了病毒感染病例。但由于各国的人口和经济情况差别较大，病毒检测能力和国家防疫政策都不尽相同，所以报告的病例是否就真实反映了病毒传播的情况?

部分专家认为鉴于无症状感染者的呼吸道标本能检出病原核酸，但由于无咳嗽、打喷嚏等临床症状，病原排出体外引起传播的机会较确诊病例相对少一些。另外，《英格兰医学杂志》上近日有报告说，一名1感染者从未出现症状，但所释放的病毒量与出现症状的人相当。因此，也有一部分科学家猜测：一些感染者“在症状轻微或无症状时具有高度传染性”。但要强调的是，类似状况的患者规模仍不清楚。早在2月17日，中国疾控中心流行病学组在《中华流行病学杂志》上发表的大规模流调论文就提到，截至2月11日，中国疾控中心共收到国内报告病例72314例，含有889例无症状感染者，比例约占1.2%。日本一个研究小组的报告称（研究论文3月12日刊登在*Eurosurveillance* 杂志），对钻石公主号游轮上的634名新冠肺炎病例进行统计模型分析，估计无症状感染者所占比例为17.9%。张文宏团队撰文指出，以目前部分研究为例，感染新冠病毒的人群中，无症状感染者的比例大约为18%—31%。不过有些患者仅出现很轻微的症状，在隔离观察期间也不一定会被发现，也常常被认为是无症状。无症状感染者的识别具有一定的困难，如何快速地、准确地、最小成本地识别和判断也是世界各国非常关注的问题。

1.2 问题的提出

2020 年 3 月 12 日，世界卫生组织(WHO)宣布，席卷全球的冠状病毒引发的病毒性肺炎(COVID-19)是一种大流行病；为此，我们需要建立合适的数学模型，分析并解决以下问题：

① 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。请建立数学模型，选择三个国家进行研究，评估它们出现第二次高峰的风险大小，并给出复工复产的政策性建议，以避免第二次高峰的出现。

② 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。例如潜伏期的长度分布，无症状感染者的比例，通常的测试方法对潜伏期和无症状感染者的假阴性率和假阳性率等等。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。请建立合理的数学模型，讨论哪些参数是最重要的，而这些参数如果不准确，会对防疫工作和疾病流行的过程带来怎样的影响。

③ 我国的无症状感染者的数量持续降低，但是并未清零，也有一些无症状感染者未被发现，请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 为了能够顺利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、模型假设

① 假设附件中提供的数据及使用的数据都真实准确，数据量化合理有效。

② 假设在神经网络预测中，输入变量作为网络的第一层合理有效。

③ 假设抽样过程中被调查人员如实填写基本信息。

④ 假设抽取实验期间没有较大人口迁入与迁出。

⑤ 假设期间没有较大死亡率与出生率。

⑥ 假设感染者能在发病的第一时间到医院就诊确认。

⑦ 假设患者的出院的时间在床位不紧张和医疗力量和资源充足的情况下。

① W 为联接权；b 为阈值。利用 Bootstrap 方法重采样，随机产生 T 个训练集 $S_1,S_2,\ldots,S_T$；

② 利用每个训练集，生成对应的决策树 C_1，$C_2\ldots$，C_T；

③ 对于测试集样本 X，利用每个决策树进行测试，得到对应的类别 $C_1(X),C_2(X)\ldots,C_T(X)$；

⑧

三、符合说明

符号	说明	符号	说明
m	时间序列	$\boldsymbol{w}$	联接权
y(t)	疫情历史值	x(t)	年份历史值
C_n	决策树	b	阈值
S_1	训练集	$\boldsymbol{e}$	误差

四、问题分析

问题一分析：在问题一中，题目要求选择三个国家进行研究，评估它们出现第二次高峰的风险大小，并给出复工复产的政策性建议，以避免第二次高峰的出现。

本文利用现有的新型冠状病毒肺炎(COVID-19)数据，选取了中国、加拿大和澳大利亚三个国家，利用 NARX 神经网络预测了三个国家未来每日新增病例，来评估三国出现第二次高峰的风险程度，并给出了复工复产的政策性建议。

问题二分析：在问题二中，题目要求请建立合理的数学模型，讨论哪些参数是最重要的，而这些参数如果不准确，会对防疫工作和疾病流行的过程带来怎样的影响。

本文通过查阅相关文献，确定新型冠状病毒影响指标；通过利用随机森林算法，求取影响因子的特征值，并对其进行排序。讨论其中影响因子最高的指标，通过分析其作用，为防疫工作献计献策。

问题三分析：在问题三中，题目要求请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。

本文结合第一阶段的模型及前两个问题的分析，评估了我国重启大型体育赛事的可能性，为设计了分阶段重启的时间表。

问题四分析：在问题四中，题目要求为了能够顺利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。

针对问题四：结合前三问的分析，为重启一些大型体育赛事，本文设计了疫情防控备忘录。

五、问题一模型的建立与求解

随着全球新型疫情拐点的到来，各国都在启动全面复工、复产的计划。为了促进企业复工复产，并对疫情出现第二次高峰的风险进行评估。本文利用现有的新型冠状病毒肺炎(COVID-19)数据，选取了中国、加拿大和澳大利亚三个国家，通过对三国每日新增病例、累计确诊病例和新增病例的增速这三组数据进行分析，来分析三国疫情的未来走势，并且选取每日新增病例作为 NARX 神经网络的输入值，利用 NARX 神经网络预测了三个国家未来每日新增病例，来评估三国出现第二次高峰的风险程度，并给出了复工复产的政策性建议。

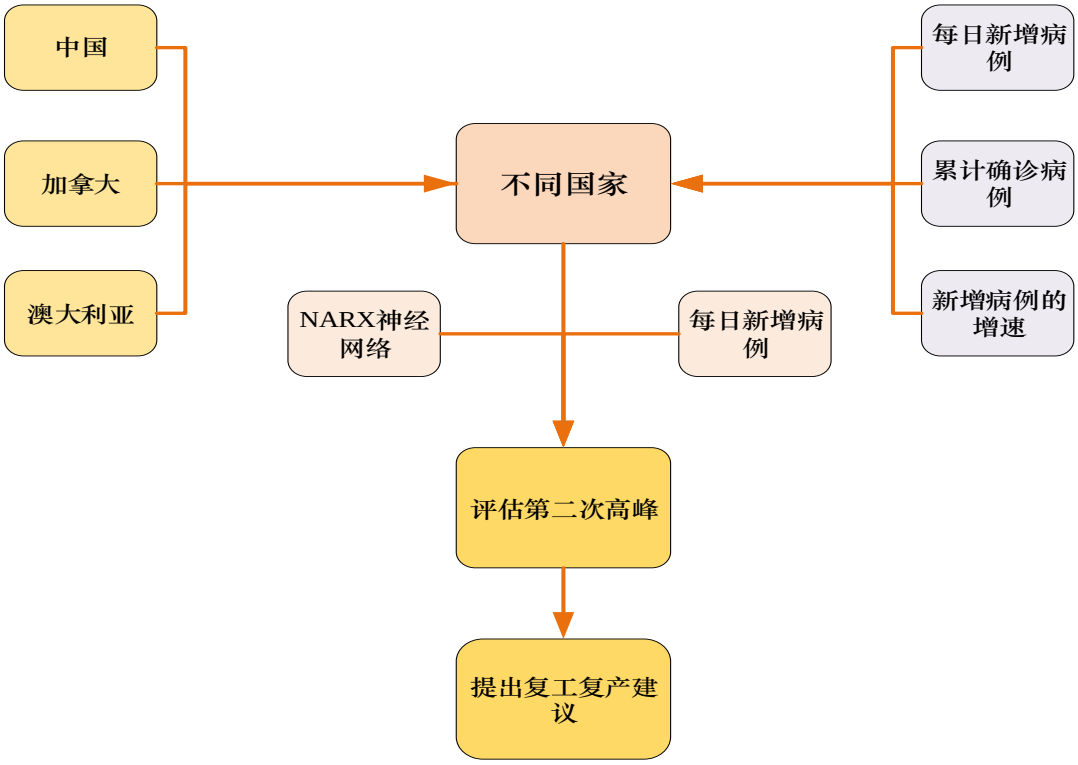

图 1.问题一求解流程图

Fig 1. Flow chart of problem 1 solution

5.1 三个国家的疫情发展现状及其特点

本文选取 2020 年 4 月 11 日至 5 月 15 日的疫情数据，从中国、加拿大和澳大利亚三个国家的疫情数据进行分析(数据来源于 https://github.com/datasets/covid-19)，新冠肺炎每日确诊病例具有一定波动，加拿大每日新增确诊病例的波动很大，但总体上呈现下降趋势，且累计确诊病例上升趋势均较为明显(见图 2、图 3)。可观测到 5 月 15 日每日新增确诊病例数有明显下降，经推测与病毒检测试剂盒的充足供给和防疫政策的实施等因素有关。而中国和澳大利亚每日新增确诊病例波动很小，且从 4 月 21 日以后，波动总体上不变，推测三国疫情已经达到拐点附进。

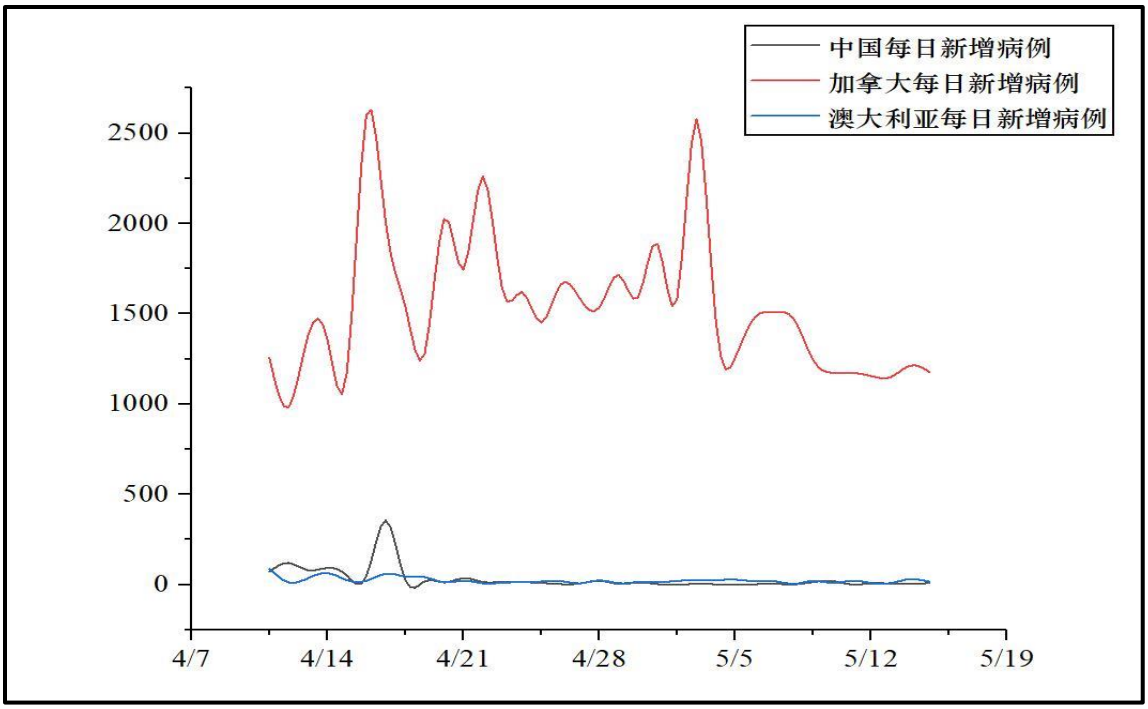

图 2 每日新增病例

Fig2. New medical records daily

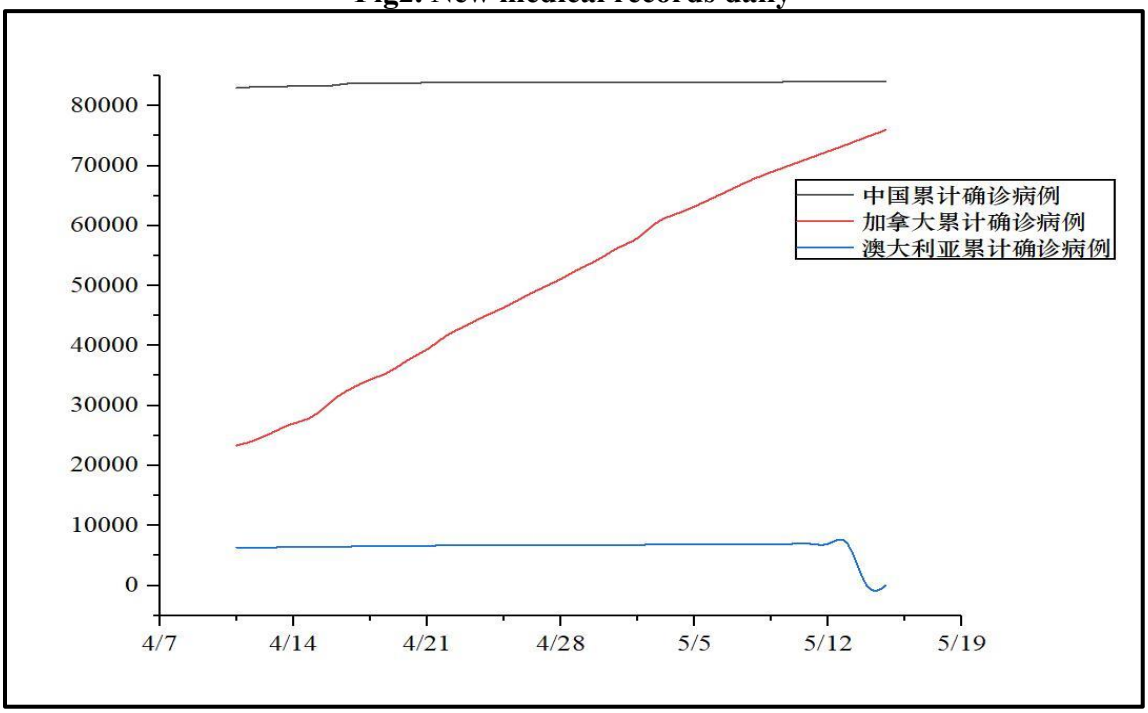

图 3 累计确诊病例

Fig3. Cumulative confirmed medical records

进一步地，从新增病例的增速来看(见图 4)，近几日来，三个国家的新增确诊病例增速均呈现下降趋势，表明在这三国政府相应防疫政策措施下，三国总体疫情正

在得到有效控制，且中国的新增病例增速基本处于 0，表明在中国政府及中国人民的努力下，疫情得到了有效的控制。并且新增确诊病例增速均呈现下降趋势，推测三国疫情已经达到拐点附进。

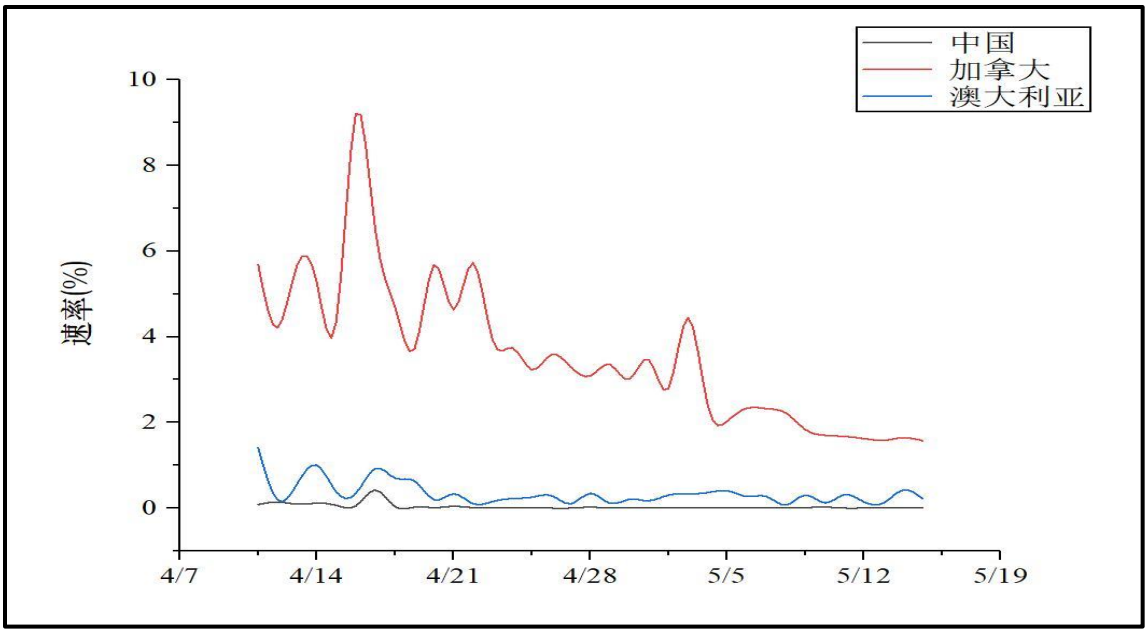

图 4 新增病例的增速

Fig5. The number of confirmed cases is increasing

5.2 NARX 神经网络原理

NARX 动态神经网络是非线性动态神经网络中应用最广泛的一种神经网络，其性能优于 BP 神经网络，模型定量预测精度较高，原理简单，计算快捷[1]。因此，选用 NARX 动态神经网络进行降雨量预测。NARX 动态神经网络模型是利用小波分析的多分辨率功能和人工神经网络的非线性逼近能力[2]。

NARX 动态神经网络同时引入 2 个时间序列，即运用被预测时间序列 y(t)的历史值和另外一个时间序列 x(t)的历史值来预测时间序列 y(t)的未来值[3]。这种形式的预测被称为具有外部输入的非线性回归，其表达式为:

$$y(t) = f(x(t-1), \dots, x(t-m), y(t-1), \dots, y(t-n)) \quad (1)$$

其中 m 为时间序列 x(t)的延时阶数。

本文中 y(t)所代表疫情数据的历史值，x(t)所代表年份的历史值；从而构成了 NARX 动态神经网络，它主要包含输入层，隐含层，输出层，输入延时，输入延时，阈值和权重构成[4]。其主要结构如图 6 所示。

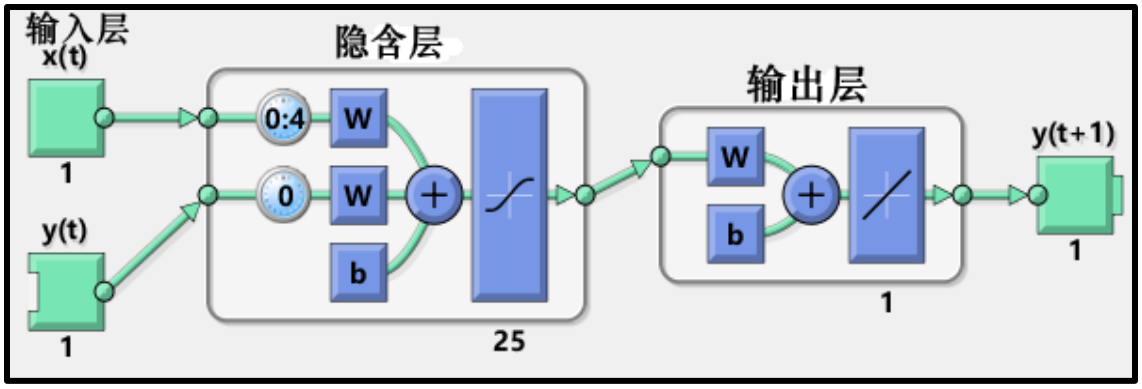

图 6 NARX 神经网络结构图

Fig.6 NARX Neural Network Structure Diagram

图 6 中，$x(t)$为神经网络的外部数据输入；$y(t)$为神经网络的数据输出；0:4 和 0 为延时阶数；W 为联接权；b 为阈值。

5.2.1 建立预测模型

动态神经网络预测模型中的参数主要包括输入延迟、输出延迟、隐层神经元个数、训练集比例、验证集比例和测试集比例。本文设置 NARX 神经网络预测模型中的参数如表 1 所示。

表 1 NARX 神经网络预测模型参数设置

Table 1 Parameters in NARX neural network prediction model

输入延迟	输出延迟	隐层神经元个数	训练集比例	验证集比例	测试集比例
1:5	1:5	25	75%	15%	15%

以拉萨市历年降雨量的数据作为 NARX 神经网络中的目标值，利用 NARX 动态神经网络进行时间序列预测，预测训练过程如图 7-图 10 所示：

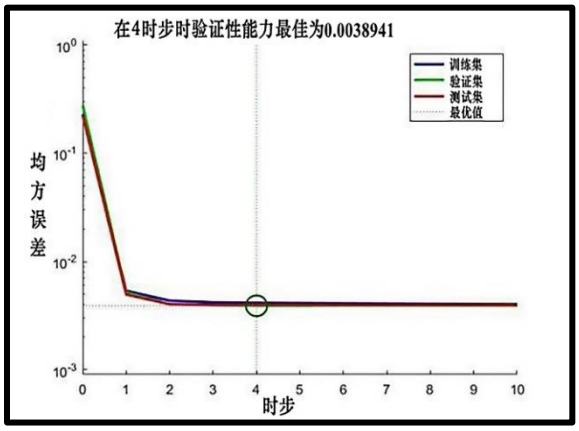

图 7 NARX 神经网络训练图

Fig.7 Training of NARX neural network

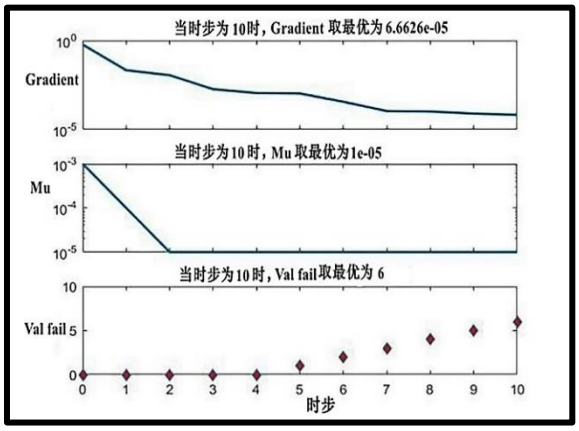

图 8 NARX 神经网络参数变化图

Fig. 8 Variation parameters of NARX neural network

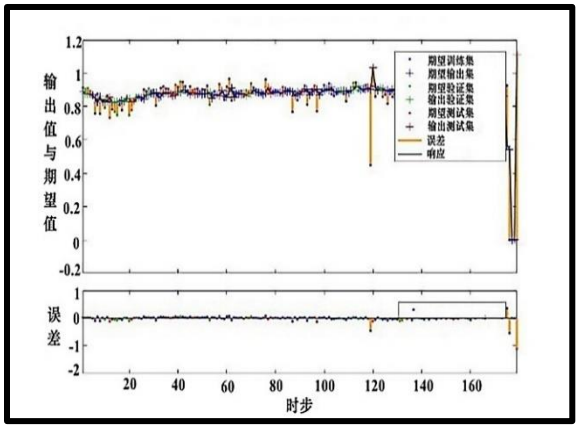

图 9 预测效果误差图

Fig.9 The error of prediction effect of neural network

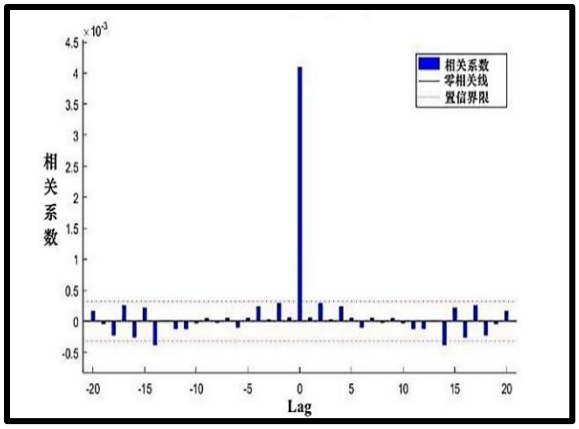

图 10 误差自相关图

Fig.10 Autocorrelation of error

NARX 神经网络训练效果如图 7 所示。从图 7 中可以看出，NARX 神经网络在训练 4 时步后训练集、验证集和测试集的均方误差基本不变并均取最优值，证明模型在第 4 时步完成自我优化，并开始样本训练，整个数据的误差此时为 0.39%。NARX 在训练过程中的梯度等参数变化如图 8(Gradient、Mu 和 Val fail 为模型训练性能的参数)。从图 8 中可以看出，模型在第 10 时步时，Gradient、Mu 和 Val fail 均达到最优值，结合图 9，证明模型从第 4 时步开始训练到第 10 时步结束。NARX 神经网络预测效果可以通过误差图 10、误差自相关图 11 进行可视化。图 9 中误差线幅度越小并且期望与输出的误差在 0 处波动越小，表示 NARX 神经网络预测效果越好；但在时步 119 处，出现个别误差线幅度过大的案例，其原因可能为原始资料数据出现偏差，但未影响模型总体误差。图 10 中误差在 lag 为 0 时应该最大，其他情况以不超过置信区间为最佳。

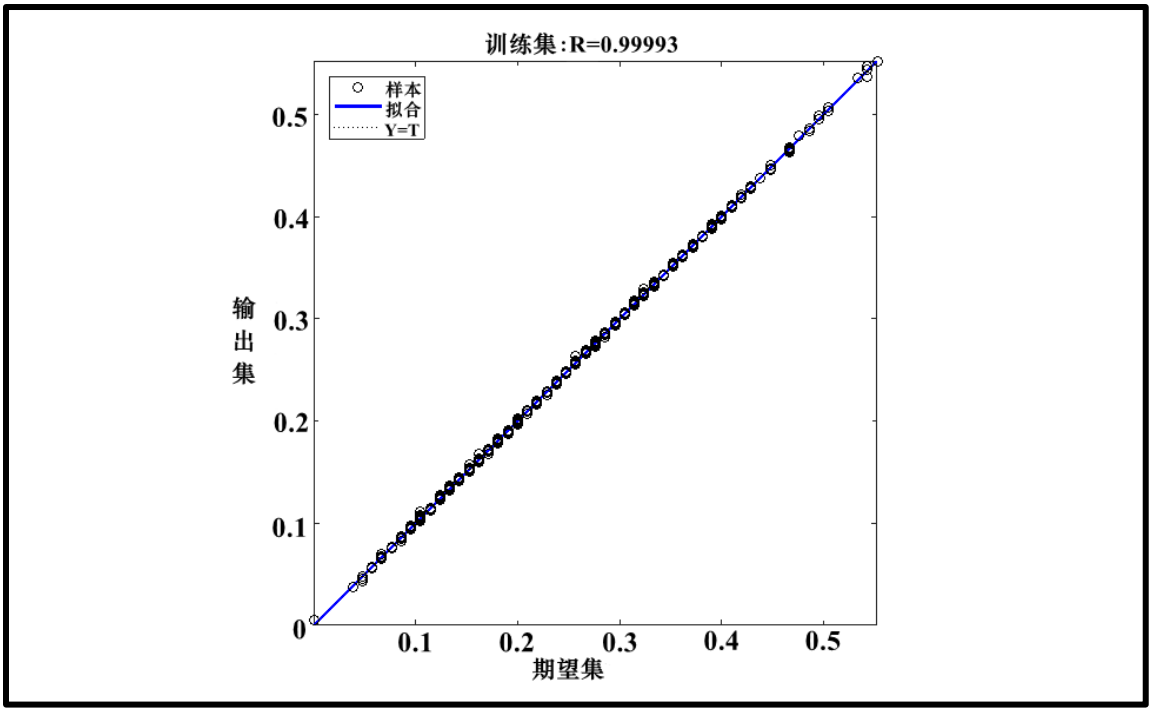

图 11 回归效果图

Fig.11 Regression effect

从 NARX 动态神经网络的回归效果图(图 11)可以看出，在模型回归训练中，样本的期望值与输出值呈线性相关，相关系数 R 为 0.99993，样本基本处于拟合线上且回归线(Y=T)与拟合线重合，表明模型拟合效果好。从而说明 NARX 神经网络的预测建模完全满足预测要求。

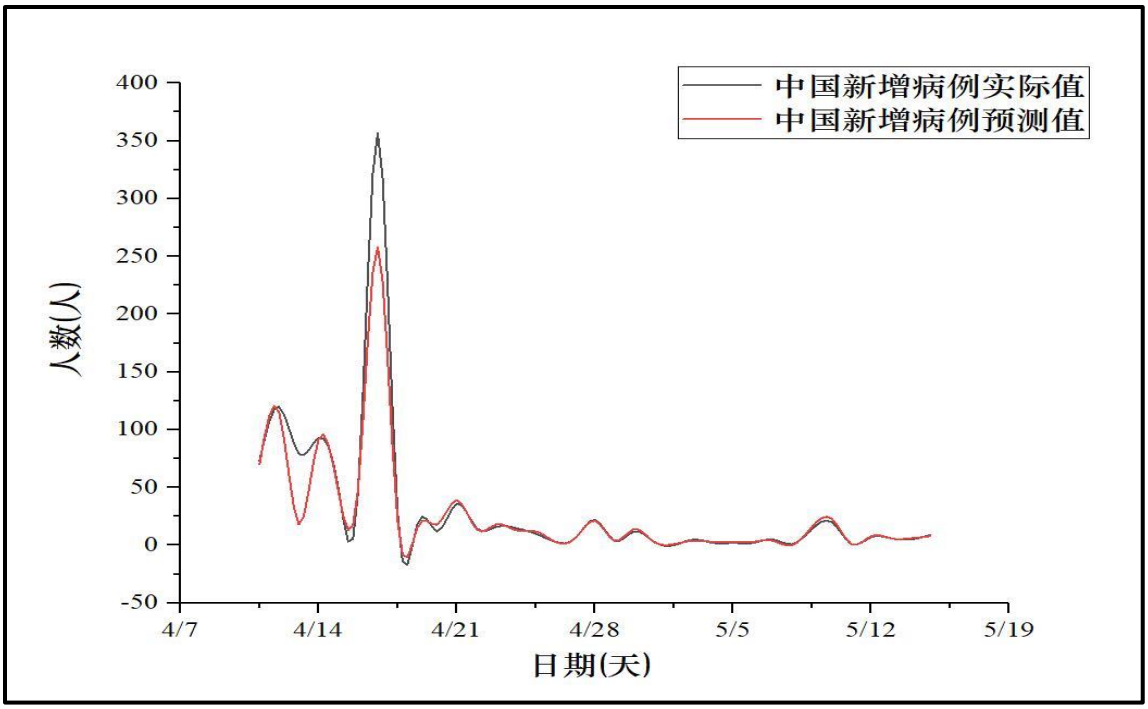

图 12 模型的预测效果图

Fig.12 The model's predictive renderings

5.2.2 预测结果检验

预测模型是否可靠，需要利用该区域没有训练的降雨量进行回判，结合训练模型数据来源，本文对中国 2020 年 2 月 7 日-2 月 19 日的新增确诊患者数据进行回判分析(表 2)。

表 2 中国新增确诊患者预测回判与误差分析

Table 2 Prediction and error analysis of newly diagnosed patients in China

日 期	7 号	8 号	9 号	10 号	11 号	12 号	13 号	14 号	15 号	16 号	17 号	18 号
预测值/mm	3520	2650	2990	2500	2100	1800	5601	6400	2012	2156	2018	1800
实测值/mm	3523	2704	3015	2525	2032	373	15136	6463	2055	2100	1921	1777
误差/%	-0.09	-2.00	-0.83	-0.92	-0.99	3.83	-63.00	-0.97	-2.09	2.67	4.63	1.29

从表 2 中可以看出，新增确诊华中预测值总误差绝对值为 1.85%，日最大误差绝对值为 63%，最小误差绝对值为 0.09%，其中 13 号的误差绝对值超过 5%，通过原因分析，由于 13 号为国家卫健委加强了对新型冠状病毒的检测标准，导致当夜武汉突增 1 万多例新型冠状病毒感染者，所以导致模型预测不准确，该因素为特殊情况。

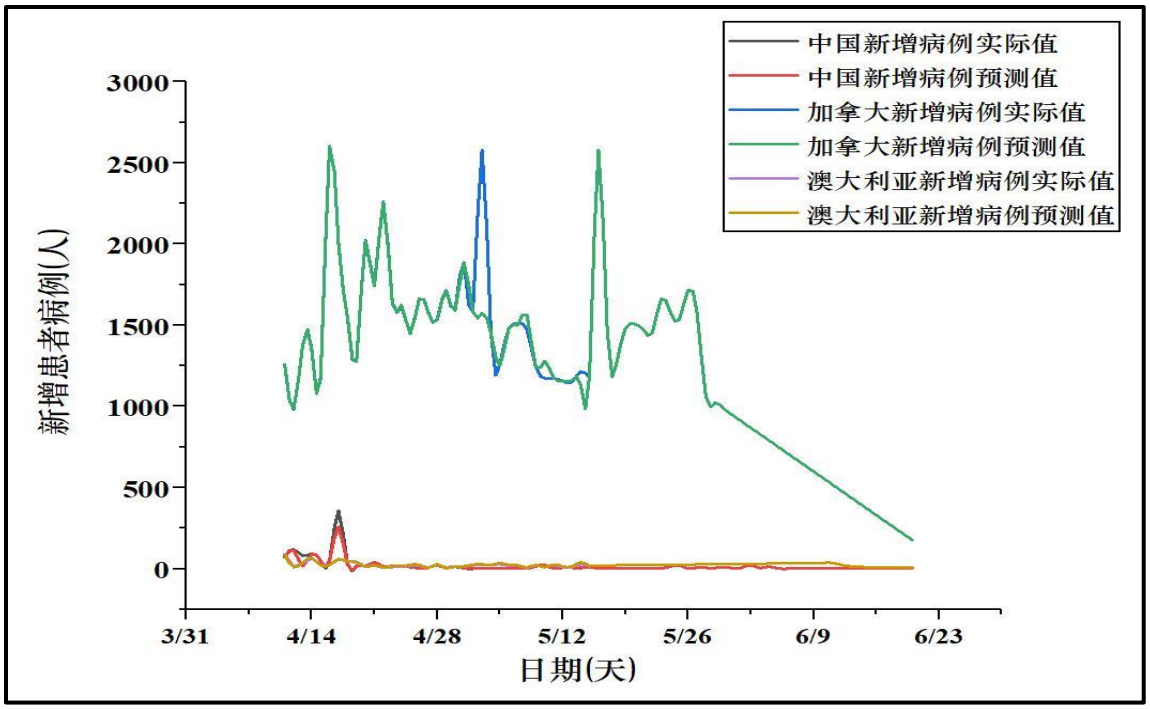

图 13 NARX 神经网络预测结果

Fig.13 NARX neural network predicted the results

从 NARX 神经网络预测模型中并结合对三国每日新增病例、累计确诊病例及确诊病例增速分析，可以得出三国政府的防疫政策的实施，使的疫情得到了有效的控制，并且通过 NARX 神经网络预测，发现三国的新增病例在未来呈现出下降的趋势，并且未出现疫情反扑的趋势。可以确定三国出现第二次高峰的风险较低，有利于复工复产。

六、问题二分析与模型构建

对于一种新出现的流行病，存在许多不确定性，例如无症状感染者比例，潜伏期的长度分布等，导致最终所得数据也存在一定的模糊性，故为了对数据进行处理，本文决定建立随机森林网络模型，以对各参数进行讨论并得出各参数的重要程度，并讨论：若参数不准确，会对防疫工作和疾病流行的过程带来怎样的影响。

如何有效地从高纬数据中提取或选择出有用的特征信息或规律，并将其分类识别已成为当今信息科学与技术所面临的基本问题[5]。本文意在建立合理的数学模型，达到对各因素完成分析的目的，因此，本文将面对的参数皆属于高维数据，故本文决定建立随机森林网络模型。

6.1 随机森林原理

随机森林算法[6]由 Leo Breiam 和 Adele Cutler 提出，该算法结合了 Breimans 的“Boot-strap aggregating”思想和 Ho 的“random subspace”方法。其实质是一个包含多个决策树的分类器，这些决策树的形式采用了随机的方法，因此也叫随机决策树，随机森林中的树之间是没有关联的。当测试数据进入随机森林时，其实就是让每一颗决策树进行分类，最后取所有决策树中分类结果最多的那类为最终结果。因此随机森林是一个包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定。

(1)Bootstrap 法重采样

设集合 S 中含有 n 个不同的样本{x_1，x_2，…….，x_n}，若每次从有放回的集合 S 中抽取一个样本，一共抽取 n 次，形成新的集合 S*，则集合 S*中不包含某个样本 x_i(i=1，2，…，n)的概率为

$$\boldsymbol{p=\left(1-\frac{1}{n}\right)^{n}} \tag{1}$$

当$n \to \infty$时，有

$$\lim_{n\to\infty} p=\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n}=e^{-1}\approx 0.368 \tag{2}$$

因此，虽然新集合 S^*样本总数与原集合 S 的样本总数相等（都为 n），但是新集合 S*中可能包含了重复的样本（有放回样本），若除去重复的样本，新集合 S*中仅包含了原集合 S 中约（1-0.368）x100%=63.2%的样本。

(2)Bagging 算法概述

Bagging 算法是最早的集成学习算法，其基本思路如图 14。

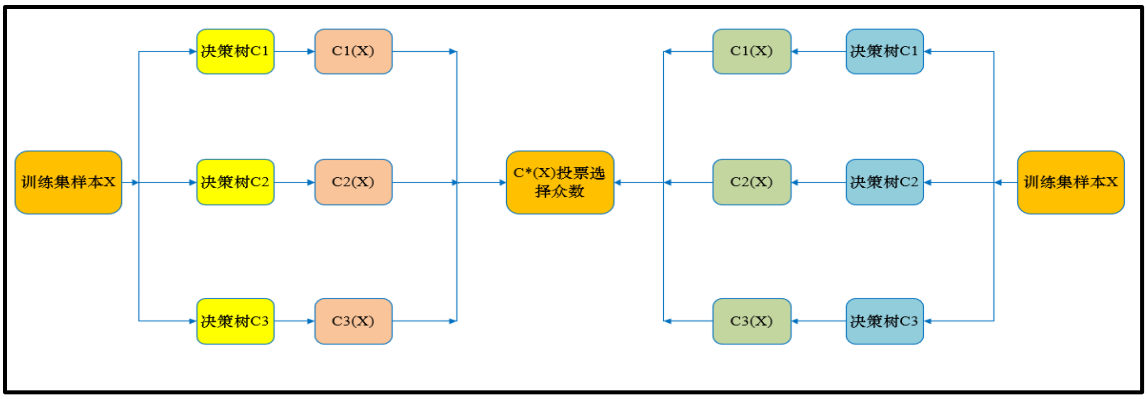

图 14 Bagging 算法基本思路
Fig14 Basic idea of Bagging algorithm

其具体的步骤可以描述为：

④ 利用 Bootstrap 方法重采样，随机产生 T 个训练集 $S_1,S_2,\ldots,S_T$；

⑤ 利用每个训练集，生成对应的决策树 C_1，$C_2\ldots$，C_T；

⑥ 对于测试集样本 X，利用每个决策树进行测试，得到对应的类别 $C_1(X),C_2(X)\ldots,C_T(X)$；

⑦ 采用投票的方法，将 T 个决策树中输出最多的类别作为测试集样本 X 所属的类别。

在确定了指标体系即模型的变量与自变量后，并不能直接将指标体系下的数据直接用于模型分析，还需要对变量数据进行分析与筛选。样本的筛选主要包括变量缺失值处理、异常值处理、变量集中度分析以及数据标准化处理。

数据处理后，还要对数据进行抽样，抽样主要包括随机和分层两种。首先采用分层抽样，进行数据抽样，目的是为了确定数据的准确性，使得模型识别好坏的能力更高。在评分建立过程前，还要将数据分为训练集，用于训练模型以及测试集，用于进行时间样本内测试。将上面分层抽样剩余样本随机抽样，抽 80%当作训练集，剩余的 20%为测试集。经上述处理最终确定了 136 组有效数据用于因素分析。

现利用随机森林神经网络对各个类型的影响因子进行分析。随机森林神经网络模型的回归效果如图 15，相关系数 R 为 0.99993，说明基于随机森林影响因素分析模型能很好的拟合和收敛数据，训练集与期望值基本上重合，说明该模型能够用于分析，可靠性较高。从图 16 模型误差图中可以看出，在随机森林中，当决策树的颗数在 50 课时，误差便大大降低，误差为 2×10^{-3}%。如果要更逼近真实值可以取 200

棵或者更多的决策树，因此本文取 200 棵进行分析。

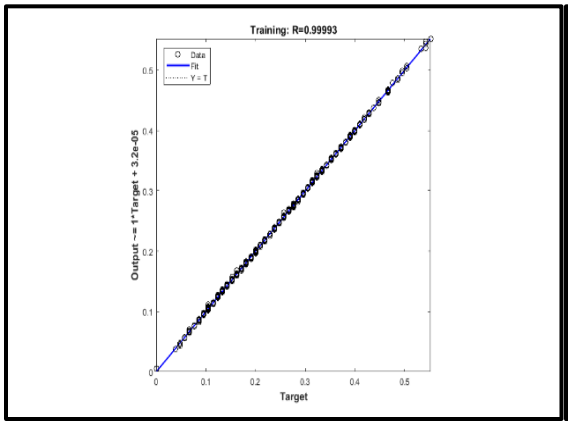

图 15 模型回归效果图

Fig15 Regression effect diagram of model

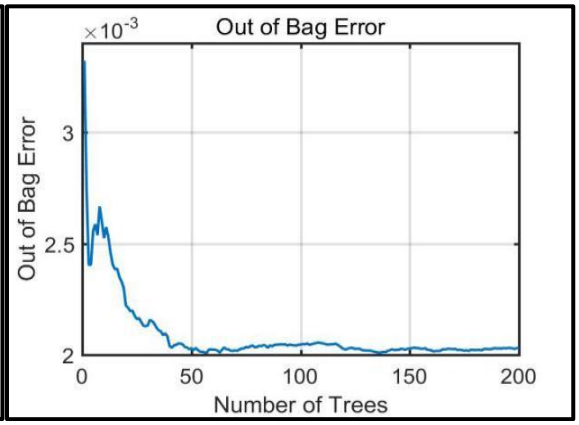

图 16 模型误差图

Fig16 Model error diagram

通过计算得出，边坡的影响因素具体结果如图 17，疫情的影响因素排序为：降感染人数、无症状感染者比例、潜伏期的长度、致死率、假阳性率、假阴性率、疫情持续时间、国家政策和疫情持续时间。

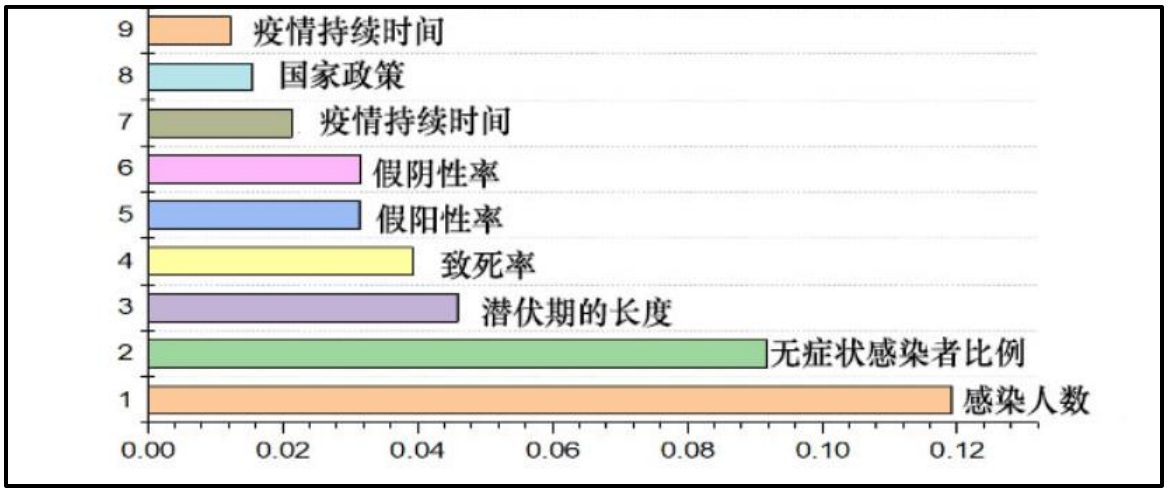

图 17 影响因素排序图

Fig17 Ranking chart of influencing factors

6.2 参数不准确的影响

从发现病患，到国务院确认为法定传染病，再到启动突发公共卫生事件应急响应的速度被大大减缓。由于是新发病毒引起的疫情，防控初期的识别难度较大，确诊能力被减弱。此阶段疫情防控以公开通报病例、组织专家研究为主，有三个方面的特点。一是新冠肺炎疫情在初期是一个风险不断积累、扩散和升级的过程，具有超出常规的传染性和传播速度加快，社会风险链已形成，但由于疫情突发，且预测

出现参数不准，在医学上尚未得到充分认识，无法明确相应的对应政策。二是疫情防控初始阶段，在“非常态”下依照常态行政程序处置，按行政层级上报和传达的速度被大大减缓。以至于无法及时进入应急响应机制。三是疫情监测机制反应不及时，新发病毒的科学认知具有一定困难，必须经过一个过程，加之新冠病毒的超常规特性增加了认知和难度，故在疫情的防控初始阶段，由于缺乏明确的诊断，参数出现了误差，只能层层人工上报，导致疫情无法及时得到处理[7]。

由于参数出现不准确，随着新冠肺炎疫情在中国快速传播蔓延，影响全国重要的交通枢纽运转，严防境外疫情输入成为当前疫情管控的重点。一是严格落实进入郑州关口的防疫措施，对海关、边检、机场、火车站等场所做好境外人员的身份识别与健康检测工作，严守疫情输入的第一道防线。二是严格实施对境外人员全面排查工作，对境外人员建立电子健康档案并实时检测，境外人员必须主动如实申报健康状况，做到有效阻断疫情传播源。三是严格采取境外人员进入郑州市后居家隔离或集中医学观察 14 天的措施，最大限度地降低疫情传播风险[8]。参数不准造成多种风险因素交织叠加的典型事件，新冠肺炎疫情目前已经在全球暴发而成为“全球性大流行病”。疫情防控是对世界各国治理体系和治理能力的重大考验，成为检验世界各国政治制度比较优势和国家治理效能的试金石。在我国抗击新冠肺炎疫情的过程中，形成了“党政主导、社会协同、公民参与”的工作格局，充分体现了制度的优势，为全球有效应对和治理疫情危机积累了宝贵的经验。

七、问题三体育赛事重启评估

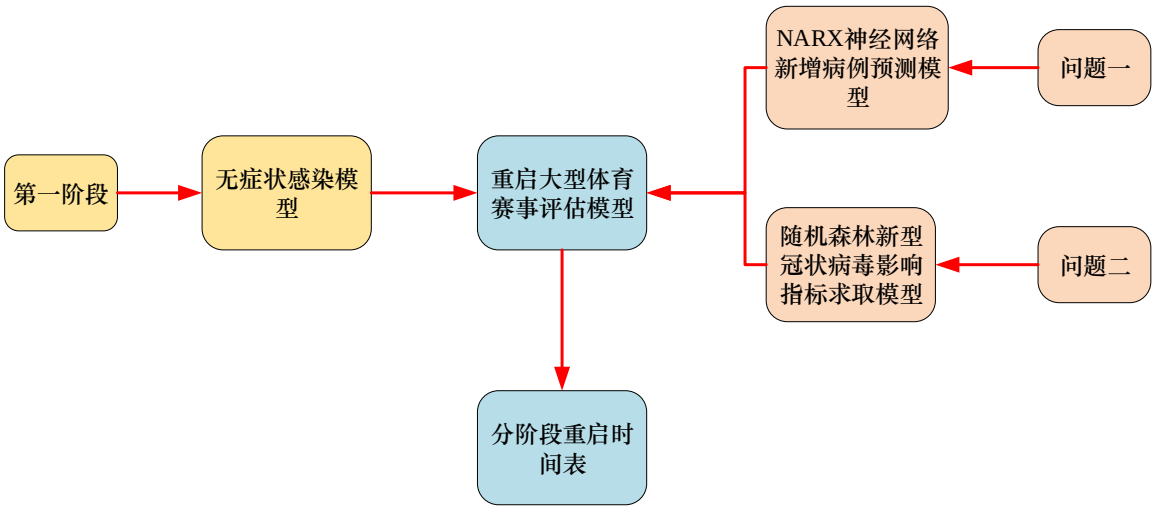

图 18 问题三的思维导图

Fig18. mind map of question 3

我国的无症状感染者的数量持续降低，但是并未清零，也有一些无症状 感染者未被发现，本文结合第一阶段的模型及前两个问题的分析，评估了我国重启大型体育赛事的可能性，为设计了分阶段重启的时间表。将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 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。

结合第一阶段构建的中国北京市的无症状感染者数量预测模型，与问题一构建的 NARX 预测新增病例模型；通过综合分析发现，中国政府实行的防疫政策对新型冠状病毒的传播起到了至关重要的作用。发现中国的疫情正在有所好转，从 NARX 神经网络预测模型中并结合对中国每日新增病例、累计确诊病例及确诊病例增速分析，可以发现中国正在想疫情的拐点附件靠近。中国新增病例在未来呈现出下降的趋势，并且未出现疫情反扑的趋势。可以确定中国出现第二次高峰的风险较低，有利于复工复产。因此可以判断出我国重启大型体育赛事的可能性很大，并且具备重启大型体育赛事的条件。

由于疫情众多大型体育赛事都被推迟，但随着疫情的减弱，国内众多疫情拟定了重启方案。再次我们以足球联赛为例。对国内大型体育赛事的重启进行预测。其中上座率的计算是用实际入场的观众人数除以整个球场的 座位数。上座率是俱乐部球迷忠诚度的标志。根据英超 1999-2000 年赛季的统计数据，曼联的上座率达到了 98.7%。由于近几年急功近利思想的蔓延，宏观调控的乏力和管理制度的缺位，我国足球职业联赛在经历了前几年高潮之后，甲 A 联赛的上座率已呈大幅度下滑趋势。

表 3 我国大型赛事重启时间表

Table 3 China's major events restart schedule

重启比赛类型 •	比赛时间
无观众赛事	5 月 20 日至六月 21 日
部分观众赛事	6 月 25 日至 7 月 25 日
全部观众但要求戴口罩赛事	8 月 11 日至 9 月 22 日
全面放开赛事	10 月 14 日以后

八、问题四的备忘录

发文人：2020 年“认证杯”数学中国数学建模参赛人

发文日期：2020 年 5 月 17 日

主题：体育赛事重启期间的疫情防控

内容：

疫情突发间，很多重大体育活动被迫推迟，但随着疫情的不断减弱，一些大型体育活动将决定重启。在此本组将给有关部门写一份有关疫情防控的备忘录：

一、赛前准备工作的内容

1.每日对准备参赛运动员两次体温检测，实行“日报告”制度，。如出现体温异常情况，第一时间上报辅导员或教练。

2.加强个人防护。运动场所内出入各场所均需佩戴口罩。应当保持运动馆、活动中心、餐厅等场所环境卫生整洁。

3.各类生活、运动、工作场所（如宿舍、体育活动场所、餐厅、洗手间等）加强通风换气。每日通风不少于 3 次，每次不少于 30 分钟。课间尽量开窗通风，保证充足的新风输入。

4.餐（饮）具应当一人一具一用，建议运动员参赛期间自带餐具。

5.宿舍要定期清洁，做好个人卫生。被褥及个人衣物要定期晾晒、定期洗涤。

6.垃圾及时收集清运，并做好垃圾盛装容器的清洁。

7.加强个人防护。去公共场所、乘坐公共交通工具、厢式电梯等必须正确佩戴口罩，做到不聚集，食堂就餐避免面对面就坐，排队前后间隔 1 米以上等等。

8.餐前、便前便后、接触垃圾、外出归来、使用体育器材等触摸眼睛等“易感”部位之前，接触污染物品之后，均要洗手。洗手时应当采用洗手液或肥皂，在流动水下按照正确洗手法彻底洗净双手，也可使用速干手消毒剂揉搓双手。

二、比赛过程防控要求

1.比赛期间减少运动员的同时参赛人数、实行多次比赛、每赛少人员的方法，每场比赛限定 300 人的入场人数，减少运动员期间的接触。

2.实行错时错峰参赛、不同比赛放到不同时间、延长竞赛时间。

3.竞赛期间禁止观赛，采用直播方式，使观众再见观赛，一次减少人员聚集，增大空气流动。

4.参赛期间裁判及评委应做好防控工作，做到戴口罩工作，且应减少裁判及评审数量。

5.竞赛期间，运动选手每周进行一次核算检测，若为阳性立即实行隔离。

6．若有国外运动选手，进行同时不同地竞赛，且由相同评委进行评判。

7.特殊时间给运动选手提供援助目的，选手将根据自己的排名收到相应的奖金。

8.一旦出现感染病例，立即将赛事中止三周或以上。

三、赛后整理防控措施

1.参赛选手在竞赛完成后进行一段时间 14 天隔离，隔离期间，运动会主办方安排服务人员每天按时安排饮食。

2.隔离期间，每天定时进行体温测量，且进行多次核酸检测。

3.隔离期间，不同国家，不同地区选手分开隔离，严禁隔离期间选手相互接触。

4.每天定时安排工作人员进行消毒杀菌工作，以保证运动员的生活卫生条件。

5.隔离结束后，安排车辆或其他交通工具吧不同地区运动员送到指定地点。

收文单位：省体育局

九、参考文献

[1]李明，杨汉生，杨成梧，等.一种改进的 NARX 回归神经网络[J].电气自动化，2006，28(4)：6-8+11.

[2]苏莉，齐勇，金玲玲，等.基于非线性多参数模型的软件老化检测[J].计算机科学，2013，40(1)：161-165+170.

[3]刘亚秋，马广富，石忠.NARX 网络在自适应逆控制动态系统辨识中的应用[J].哈尔滨工业大学学报，2005，37(2)：173-176.

[4]孙国祥，闫婷婷，汪小昂，等.基于小波变换和动态神经网络的温室黄瓜蒸腾速率预测[J].南京农业大学学报，2014，37(5)：143-152.

[5]蒋胜利．高维数据的特征选择与特征提取研究[D]．西安：西安电子科技大学计算机学院，2011.

[6] BREIMAN L. Random Forests[J].Machine Learning，2001，45（1）:5~32.

[7]ARMOUR G C, BUFFA E S. A heuristic algorithm andsimulation approach to relative location of facilities[J].Management Science, 1963, 9(2): 294-309.

[8]赵宏波,魏甲晨,王爽,刘雅馨,李光慧,苗长虹.大城市新冠肺炎疫情风险评估与精准防控对策——以郑州市为例[J].经济地理,2020,40(04):103-109+124.

[9]赵宏波,魏甲晨,王爽,刘雅馨,李光慧,苗长虹.大城市新冠肺炎疫情风险评估与精准防控对策——以郑州市为例[J].经济地理,2020,40(04):103-109+124.

附件

一、NARX 动态神经网络代码

```
% Solve an Autoregression Problem with External Input with a NARX Neural
Network
% Script generated by Neural Time Series app
%
% This script assumes these variables are defined:
%
%     P - input time series.
%     F - feedback time series.

X = tonndata(F,true,false);
T = tonndata(P,true,false);

% Choose a Training Function
% For a list of all training functions type: help nntrain
% 'trainlm' is usually fastest.
% 'trainbr' takes longer but may be better for challenging problems.
% 'trainscg' uses less memory. NTSTOOL falls back to this in low memory
situations.
trainFcn = 'trainlm';    % Levenberg-Marquardt

% Create a Nonlinear Autoregressive Network with External Input
inputDelays = 1:5;
feedbackDelays = 1;5;
hiddenLayerSize = 25;
net = narxnet(inputDelays,feedbackDelays,hiddenLayerSize,'open',trainFcn);
```

```
% Prepare the Data for Training and Simulation
% The function PREPARETS prepares timeseries data for a particular network,
% shifting time by the minimum amount to fill input states and layer states.
% Using PREPARETS allows you to keep your original time series data unchanged,
while
% easily customizing it for networks with differing numbers of delays, with
% open loop or closed loop feedback modes.
[x,xi,ai,t] = preparets(net,X,{},T);

% Setup Division of Data for Training, Validation, Testing
net.divideParam.trainRatio = 75/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;
%ÍøÂçÑµÁ·º¯ÊýÉè¶¨
net.trainFcn = 'trainlm'; % Levenberg-Marquardt

%Îó²îº¯ÊýÉè¶¨
net.performFcn = 'mse';    %Mean squared error

% Train the Network
[net,tr] = train(net,x,t,xi,ai);

% Test the Network
y = net(x,xi,ai);
e = gsubtract(t,y);
performance = perform(net,t,y);
% View the Network
```

```
% view(net)


% Plots
% Uncomment these lines to enable various plots.
%figure, plotperform(tr)
%figure, plottrainstate(tr)
%figure, plotregression(t,y)
%figure, plotresponse(t,y)
%figure, ploterrcorr(e)
%figure, plotinerrcorr(x,e)


% Step-Ahead Prediction Network
% For some applications it helps to get the prediction a timestep early.
% The original network returns predicted y(t+1) at the same time it is given y(t+1).
% For some applications such as decision making, it would help to have predicted
% y(t+1) once y(t) is available, but before the actual y(t+1) occurs.
% The network can be made to return its output a timestep early by removing one
delay
% so that its minimal tap delay is now 0 instead of 1.    The new network returns the
% same outputs as the original network, but outputs are shifted left one timestep.
nets = removedelay(net);
nets.name = [net.name ' - Predict One Step Ahead'];
view(nets)
[xs,xis,ais,ts] = preparets(nets,X,{},T);
ys = nets(xs,xis,ais);
stepAheadPerformance = perform(nets,ts,ys);
% Multi-step Prediction
% Sometimes it is useful to simulate a network in open-loop form for as
```

```
% long as there is known output data, and then switch to closed-loop form
% to perform multistep prediction while providing only the external input.
% Here all but 5 timesteps of the input series and target series are used to
% simulate the network in open-loop form, taking advantage of the higher
% accuracy that providing the target series produces:
numTimesteps = size(x,2);
knownOutputTimesteps = 1:(numTimesteps-5);
predictOutputTimesteps = (numTimesteps-4):numTimesteps;
X1 = X(:,knownOutputTimesteps);
T1 = T(:,knownOutputTimesteps);
[x1,xio,aio] = preparets(net,X1,{},T1);
[y1,xfo,afo] = net(x1,xio,aio);
% Next the the network and its final states will be converted to closed-loop
% form to make five predictions with only the five inputs provided.
x2 = X(1,predictOutputTimesteps);
[netc,xic,aic] = closeloop(net,xfo,afo);
[y2,xfc,afc] = netc(x2,xic,aic);
multiStepPerformance = perform(net,T(1,predictOutputTimesteps),y2);
% Alternate predictions can be made for different values of x2, or further
% predictions can be made by continuing simulation with additional external
% inputs and the last closed-loop states xfc and afc.


% Deployment
% Change the (false) values to (true) to enable the following code blocks.
% See the help for each generation function for more information.
if (false)
    % Generate MATLAB function for neural network for application deployment
    % in MATLAB scripts or with MATLAB Compiler and Builder tools, or simply
```

```
    % to examine the calculations your trained neural network performs.
    genFunction(net,'myNeuralNetworkFunction');
    y = myNeuralNetworkFunction(x,xi,ai);
end
if (false)
    % Generate a matrix-only MATLAB function for neural network code
    % generation with MATLAB Coder tools.
    genFunction(net,'myNeuralNetworkFunction','MatrixOnly','yes');
    x1 = cell2mat(x(1,:));
    x2 = cell2mat(x(2,:));
    xi1 = cell2mat(xi(1,:));
    xi2 = cell2mat(xi(2,:));
    y = myNeuralNetworkFunction(x1,x2,xi1,xi2);
end
if (false)
    % Generate a Simulink diagram for simulation or deployment with.
    % Simulink Coder tools.
    gensim(net);
end
```

二、随机森林代码

```
%-----------------------------------------------------------------------
% Load an example dataset provided with matlab
In = unnamed';
Out =unnamed1';


%-----------------------------------------------------------------------
% Find capabilities of computer so we can best utilize them.
```

```
% Find if gpu is present
ngpus=gpuDeviceCount;
disp([num2str(ngpus) ' GPUs found'])
if ngpus>0
    lgpu=1;
    disp('GPU found')
    useGPU='yes';
else
    lgpu=0;
    disp('No GPU found')
    useGPU='no';
end

% Find number of cores
ncores=feature('numCores');
disp([num2str(ncores) ' cores found'])

% Find number of cpus
import java.lang.*;
r=Runtime.getRuntime;
ncpus=r.availableProcessors;
disp([num2str(ncpus) ' cpus found'])

if ncpus>1
    useParallel='yes';
else
    useParallel='no';
end
```

```
[archstr,maxsize,endian]=computer;
disp([...
    'This is a ' archstr ...
    ' computer that can have up to ' num2str(maxsize) ...
    ' elements in a matlab array and uses ' endian ...
    ' byte ordering.'...
    ])

% Set up the size of the parallel pool if necessary
npool=ncores;

% Opening parallel pool
if ncpus>1
    tic
    disp('Opening parallel pool')

    % first check if there is a current pool
    poolobj=gcp('nocreate');

    % If there is no pool create one
    if isempty(poolobj)
        command=['parpool(' num2str(npool) ');'];
        disp(command);
        eval(command);
    else
        poolsize=poolobj.NumWorkers;
        disp(['A pool of ' poolsize ' workers already exists.'])
```

```
    end

    % Set parallel options
    paroptions = statset('UseParallel',true);
    toc

end

%-------------------------------------------------------------------
tic
leaf=5;
ntrees=1000;
fboot=1;
surrogate='on';
disp('Training the tree bagger')
b = TreeBagger(...
        ntrees,...
        In,Out,...
        'Method','regression',...
        'oobvarimp','on',...
        'surrogate',surrogate,...
        'minleaf',leaf,...
        'FBoot',fboot,...
        'Options',paroptions...
    );
toc

%-------------------------------------------------------------------
```

```
% Estimate Output using tree bagger
disp('Estimate Output using tree bagger')
x=Out;
y=predict(b, In);
name='Bagged Decision Trees Model';
toc

%--------------------------------------------------------------------------
% calculate the training data correlation coefficient
cct=corrcoef(x,y);
cct=cct(2,1);

%--------------------------------------------------------------------------
% Create a scatter Diagram
disp('Create a scatter Diagram')

% plot the 1:1 line
plot(x,x,'LineWidth',3);

hold on
scatter(x,y,'filled');
hold off
grid on

set(gca,'FontSize',18)
xlabel('Actual','FontSize',25)
ylabel('Estimated','FontSize',25)
title(['Training Dataset, R^2=' num2str(cct^2,2)],'FontSize',30)
```

```
drawnow

fn='ScatterDiagram';
fnpng=[fn,'.png'];
print('-dpng',fnpng);

%--------------------------------------------------------------------------
% Calculate the relative importance of the input variables
tic
disp('Sorting importance into descending order')
weights=b.OOBPermutedVarDeltaError;
[B,iranked] = sort(weights,'descend');
toc

%--------------------------------------------------------------------------
disp(['Plotting a horizontal bar graph of sorted labeled weights.'])

%--------------------------------------------------------------------------
figure
barh(weights(iranked),'g');
xlabel('Variable Importance','FontSize',30,'Interpreter','latex');
ylabel('Variable Rank','FontSize',30,'Interpreter','latex');
title(...
    ['Relative Importance of Inputs in estimating Redshift'],...
    'FontSize',17,'Interpreter','latex'...
    );
hold on
```

```
barh(weights(iranked(1:10)),'y');
barh(weights(iranked(1:5)),'r');

%-------------------------------------------------------------------------
grid on
xt = get(gca,'XTick');
xt_spacing=unique(diff(xt));
xt_spacing=xt_spacing(1);
yt = get(gca,'YTick');
ylim([0.25 length(weights)+0.75]);
xl=xlim;
xlim([0 2.5*max(weights)]);

%-------------------------------------------------------------------------
% Add text labels to each bar
for ii=1:length(weights)
    text(...
        max([0 weights(iranked(ii))+0.02*max(weights)]),ii,...
        ['Column ' num2str(iranked(ii))],'Interpreter','latex','FontSize',11);
end

%-------------------------------------------------------------------------
set(gca,'FontSize',16)
set(gca,'XTick',0:2*xt_spacing:1.1*max(xl));
set(gca,'YTick',yt);
set(gca,'TickDir','out');
set(gca, 'ydir', 'reverse' )
set(gca,'LineWidth',2);
```

```
drawnow

%-------------------------------------------------------------------------
fn='RelativeImportanceInputs';
fnpng=[fn,'.png'];
print('-dpng',fnpng);

%-------------------------------------------------------------------------
% Ploting how weights change with variable rank
disp('Ploting out of bag error versus the number of grown trees')

figure
plot(b.oobError,'LineWidth',2);
xlabel('Number of Trees','FontSize',30)
ylabel('Out of Bag Error','FontSize',30)
title('Out of Bag Error','FontSize',30)
set(gca,'FontSize',16)
set(gca,'LineWidth',2);
grid on
drawnow
fn='EroorAsFunctionOfForestSize';
fnpng=[fn,'.png'];
print('-dpng',fnpng);
```