本科毕业论文（设计）

论文题目（中文）突发传染病疫情预测模型对比分析研究——以新冠疫情为例

论文题目（英文）Comparative Analysis of Epidemic Prediction Models for Outbreaks of Infectious Diseases -- a Case Study of COVID-19

学生姓名 贾凯文

指导教师 曲宗希

学　　院 管理学院

专　　业 信息管理与信息系统

年　　级 2017 级

诚信责任书

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文(设计)中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人或集体已经发表或在网上发表的论文。

本声明的法律责任由本人承担。

论文作者签名：贾凯文　　日　期：2021.5.25

关于毕业论文（设计）使用授权的声明

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文。本人离校后发表、使用毕业论文或与该论文直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文研究内容：

□可以公开

□不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名：贾凯文　　导师签名：

日　　期：2021.5.25　　日　　期：2021.5.25

突发传染病疫情预测模型对比分析研究——以新冠疫情为例

中文摘要

新冠肺炎疫情肆虐全球，受到世界各国政府与学者的高度关注。对疫情发展态势的预测与判断对疫情防控具有重要意义。然而，如何对传染病疫情进行精准预测一直是公共卫生领域的最具挑战性的热点研究问题之一。在本文中，我们首先对以往各种传染病疫情预测模型进行分类，在每类中分别选择具有代表性的预测模型：SEIR 传染病动力学模型、ARIMA 时间序列模型和 LSTM 神经网络模型，并介绍这些方法的基本原理。之后我们分别用这三类模型对六个不同国家（德国、日本、俄罗斯、巴西、印度和南非）三个时期的新冠肺炎疫情发展态势进行预测，选择合适的指标对他们的预测结果进行评价，对比他们的预测效果，并讨论各类模型的适用性，最后提出可行的疫情预测策略，为疫情防控提供科学依据。

关键词：新冠肺炎疫情；传染病预测模型；SEIR 模型；ARIMA 模型；LSTM 模型

Comparative Analysis of Epidemic Prediction Models for Outbreaks of Infectious Diseases -- a Case Study of COVID-19

Abstract

The COVID-19 epidemic has been spread around the world and caused a lot of damage. Many scientists try to explore how will the epidemic develop. It's important for epidemic prevention and control to predict the trend of epidemic. However, in the field of sanitary science, precisely forecasting where the epidemic heading is the hottest topic. In this paper, we first classify the previous epidemic prediction models of infectious diseases, select the representative prediction models of each category: SEIR infectious disease dynamics model, ARIMA time series model and LSTM neural network model. We introduce their rationale and use them to forecast the epidemic trend in six different countries (Germany, Japan, Russia, Brazil, India and South Africa). in order to compare the predicted results better, we divide the time into three segments. Then we choose the appropriate indicators to measure which one is going to be better, and discuss which cases the models fit. At last, based on the results of comparison and evaluation, we proposed reasonable epidemic prevention and control strategies.

Keywords: COVID-19; Infectious disease prediction model; SEIR model; ARIMA model; LSTM model;

目 录

第一章 绪论

1.1 研究背景与意义

2019 年 12 月底，在中国武汉相继出现不明原因肺炎病例，其传播速度之快超出预计，很快在中国国内引起大规模爆发，引起该肺炎的元凶是一种从未发现过的病毒，世界卫生组织将其命名为 2019 冠状病毒病(corona virus disease 2019，COVID-19)。该病毒致病性强，且传播十分迅速，疫情在中国爆发时正处于春运高峰，大范围的人口流动导致了大规模的疫情爆发，其感染人数达到非典感染人数的十倍以上，新冠肺炎疫情已成为严重的公共卫生问题。为遏制疫情进一步恶化，全国政府纷纷出台防控政策，通过隔离病患和疑似患者，限制人口流动和出行等方式，有效降低了病毒感染率，使疫情得到控制。我国将新冠肺炎病毒病纳入乙类传染病，并采取甲类防控措施。截至 2021 年 2 月 28 日，我国累计确诊病例 101,920 例，累计死亡 4,844 例；全球累计确诊病例 114,395,350 例，累计死亡 2,534,727 例。

新冠肺炎疫情受到了各个国家的高度重视，准确的分析和预测可以帮助政府在疫情早期做出正确的判断，对政府制定合理的防控策略具有重要意义。余艳妮[1]等对传染病预测模型进行了分类，将模型划分为传统回归方法，传染病动力学模型以及多元统计三类。朱仁杰[2]等利用 SIR 模型对 7 个疫情较为严重的国家进行预测，结果表明该模型用于新冠肺炎疫情的预测是可靠的。Jia wenxiao[3]等利用人工神经网络模型和时间序列模型对中国十种常见传染病进行了预测，得出人工神经网络模型对这些疾病有较好的预测能力。温亮[4]等使用时间序列模型对巴基斯坦疫情发展趋势做出预测，在短期预测中误差低于 5%。Dehning, J [5]等关注德国的疫情发展态势，通过建立贝叶斯流行传染病学模型，可以帮助预测干预措施实施后疫情的变化趋势。周涛[6]等对 SEIR 模型中的基本再生数参数做出初步估计，得出新冠肺炎病毒属于中高度传染性的疾病。杨雨琦[7]等通过建立 SIR 传染病模型预测重庆市的疫情发展态势。梅文娟[8]等改进了 SIR 模型，引入了机器学习方法，提出了极限 IR 模型，可准确实现疫情的实时预测。

综上所述，学者们聚焦于利用单一模型对新冠肺炎疫情进行预测，其中 SEIR 模型被使用得最为频繁。一些学者对传染病预测模型进行了总结，但是缺乏定量研究，停留在综述层面。很少有学者对新冠肺炎疫情预测模型的预测效果进行比较。

1.2 研究目的

学者们进行了许多关于新冠肺炎疫情预测的研究，这些研究是基于传染病疫情预测模型，结合新冠肺炎病毒的特性来对疫情发展态势做出预测和判断，但绝大多数研究是用单一方法预测，很少有研究者对不同模型之间的预测效果进行比较。本文首先通过对传染病预测模型进行分类，从不同类别的预测模型中选择最具有代表性的模型，对六个不同国家不同时期的疫情发展态势做出拟合，比较模型的预测效果，讨论他们的适用性并给出合理可行的疫情预测方案。

第二章 传染病模型分类

对于传染病的预测和分析是公共卫生领域的重要课题，传染病预测方法可以分为定性预测和定量预测，我们这里主要对定量预测模型进行分类。定量预测模型是利用数学模型，根据传染病的各类特征来确定参数，模拟传染病传播过程的一类模型，其预测的准确程度主要依赖于参数对于传染病特征的刻画情况，参数越精准，模型的拟合效果就越好[9]。学者们通过建立各类不同的模型对传染病的发展趋势做出拟合与预测，传染病预测模型可以分为三类：传染病动力学模型，时间序列模型及人工神经网络模型。

传染病动力学模型是一种微分方程模型，描述的是封闭状态下的疫情传播，即人口无迁入迁出，不考虑任何其他的干预措施[10]。经典的动力学模型包括 SI、SIR、SEIR 等，其中 SEIR 模型与新冠肺炎疫情的传播模式较为吻合，被广泛应用到疫情预测中[11]。学者们根据疫情实际情况提出了很多改进的 SEIR 模型用于预测新冠肺炎疫情，来对疫情拐点和发展做出预测，在一定程度上对疫情防控起到了指导作用[12]。

时间序列模型也常被用作传染病疫情预测，该方法认为病例的增长人数是历史数据的加权和，假设疫情的发展仅与时间有关，根据过去疫情的情况来推测其未来的发展，主要被用于疫情短期和中期的预测。王旭艳等[13]利用指数平滑模型对湖北省疫情进行分析预测。杨真真[14]等利用 ARIMA 时间序列模型对美国疫情发展做出拟合和预测，取得了较好的效果。时间序列预测模型的优势在于所需要的信息较少，克服了动力学模型所需要的精确参数不易得到的缺陷，在疫情初期信息不透明的情况下，可以使用该模型对疫情做出预测分析。

随着各种神经网络算法的迅速发展，神经网络在预测新冠肺炎疫情上发挥了重要作用，越来越多的被用在传染病的分析和预测中。吴志强[15]等针对传统模型的缺陷，使用组合神经网络对疫情出传播做出预测。Tomar,A[16]等采用 LSTM 方法预测印度疫情的发展情况，并验证了封锁和隔离等防疫举措的效果。王志心等采用数学建模的方式，通过机器学习对新型冠状病毒肺炎确诊人数趋势进行了预测，成功预判了疫情走向，并对中国各省累计确

诊病例所占的比例进行了预测[17]。人工神经网络模型可以适用于任何场景，学习能力强，相比传统的模型，它的预测效果也很好。神经网络模型既不需要获取参数，对数据的形式也无任何要求，打破了传统模型需要让数据适应自身的现状，因为它的这些优点，人工神经网络模型成为了疫情预测的热点模型之一，在疫情预测中的应用越来越广泛。

为了对比三种不同类型模型的优缺点，探索他们的适用情况，本文拟通过建立 SEIR 传染病动力学模型，ARIMA 时间序列模型，LSTM 神经网络模型分别对 6 个国家的疫情做出分析预测，对比模型的预测效果，并建立合理的指标评价体系对各类模型的预测结果进行评价，提出可行的疫情预测策略。

第三章 数据来源

本研究采用 6 个国家(日本、俄罗斯、印度、德国、巴西、南非)累计确诊病例数据建模。这 6 个国家的疫情数据没有出现错乱的情况；他们同属于疫情较为严重的国家，对这些国家的疫情预测不仅可以更好地检验模型的预测效果，而且更具有现实意义。

本文数据来源于世界卫生组织（WHO）公布的 COVID-19 累计确诊病例，数据的时间跨度为 2020 年 1 月 22 日至 2021 年 3 月 5 日，共计 409 天 。根据 6 个国家疫情的发展情况，我们将全部数据划分为三个阶段：2020 年 1 月 22 日至 6 月 5 日是 6 个国家第一波疫情爆发，这一阶段感染人数增长迅速，我们把这一阶段划为疫情初期；2020 年 6 月 6 日至 10 月 19 日 6 个国家接受第二波疫情爆发的考验，我们将其划分为疫情中期；在两波疫情结束后，6 个国家国内疫情走向了不同的发展道路，我们把 2020 年 10 月 19 日至 2021 年 3 月 5 日划分为疫情后期。将疫情数据划分有助于我们更好的观察各类模型在疫情不同阶段的预测表现。

第四章 模型介绍

4.1 SEIR 模型

SEIR 模型（Susceptible Exposed Infected Recovered Model），是一种经典的传染病预测模型，相较于其他模型来说，SEIR 考虑到了病毒的潜伏期，即普通人在感染病毒之后，不会立即转化成确诊病例，而是成为病毒携带者。模型把人群分为四类，易感者(susceptible，S)、感染者(infected，I)、接触者(exposed，E)和康复人群(recovered，R)。经典 SEIR 模型有四个基本假设：一、将所有人群都划分到易感者（S）人群中；二、不存在人口流动；三、病毒存在潜伏期，潜伏期内具有传染性；四、康复人群（R）不会向感染者（I）转化。建

立微分方程如下：

$$\frac{dS}{dt} = -\frac{r\beta IS}{N}（1）$$

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \frac{r\beta IS}{N} - \alpha E（2）$$

$$\frac{dI}{dt} = \alpha E - \gamma I（3）$$

$$\frac{dR}{dt} = \gamma R(4)$$

r 表示每人接触到的人数，N 表示全国总人口，r/N 即表示人群的接触率。β 表示易感者 S 被潜伏者 I 感染的概率，α 表示潜伏者向感染者转化的概率（潜伏期的倒数），这里我们设置为 1/14。γ 表示康复概率。为了更好比较三个模型的预测结果，我们引入死亡病例（D），死亡率设置为全球疫情整体死亡率 2.2%，引入累计确诊人数（confirmed，C），C=I+R+D。

4.2 ARIMA 模型

ARIMA 模型（Autoregressive Integrated Moving Average model），差分整合移动平均自回归模型，是时间序列分析预测方法之一，常被用于传染病疫情预测[18]。该模型表示为 ARIMA(p,d,q)，是 ARMA 模型的扩展，用来对非平稳序列数据进行预测分析。可以表示为：

$$\left(1 - \sum_{i=1}^{p} \phi_i L^i\right)(1-L)^d x_t = \left(1 + \sum_{i=1}^{q} \theta_i L^i\right)\varepsilon_t \ (5)$$

其中 L 是滞后算子，p 是自回归部分阶数，Φ 和 θ 是待估计参数，q 是移动平均部分阶数，d 表示为了使数据成为平稳数据而进行的差分次数[19]，在本研究中，6 个国家的累计确诊病例数据经过二阶差分后转换为平稳序列，d 确定为 2。p 和 q 的值我们通过 AIC 准则来确定[20]：

$$AIC = -2\,log(L) + 2(p+q+k)（6）$$

L 是似然函数，k 是 ARIMA 模型的截距。AIC 准则会帮助我们选择出拟合程度较好且不会出现过拟合情况的模型参数，我们选择 AIC 值最小的情况来确定 p 和 q 的值，在本研究中，ARIMA（3，2，3）被认为是较为合适的模型参数。

4.3 LSTM 模型

LSTM（Long Short-Term Memory）长短期记忆网络，是一种时间循环神经网络，是循环神经网络的改进模型，经典循环神经网络包括三部分：输入层，隐含层和输出层，在隐含层中，经过计算的输入值可以被输入到下一时间的隐藏层中，其基本结构如下图所示：

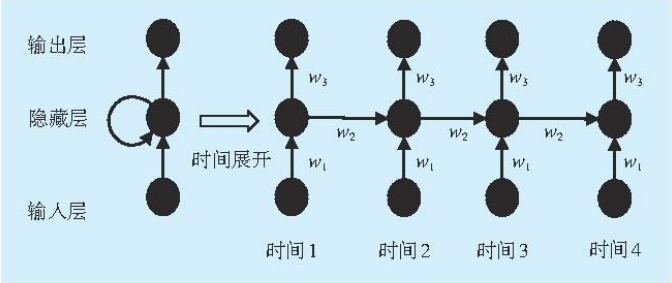

图 4.1 RNN 模型基本单元[21]

RNN 模型存在一个比较明显的缺陷，即在时间间隔较长的情况下，模型可能会失去先前的输入信息。LSTM 模型通过改进隐含层的结构克服了这一点，LSTM 模型的隐含层中包含了输入门，遗忘门，记忆细胞和输出门。信息在被输入到隐含层后，遗忘门会筛去不需要储存的信息，其余信息储存在记忆细胞中，最后的输出是输出门根据细胞状态、输入信息和上一时刻的输出结果计算得出的。本文建立了一个包含了 96*3 个隐藏单元的 LSTM 神经网络，模型的 dropout（丢弃概率）设置为 0.5，避免出现过拟合情况。

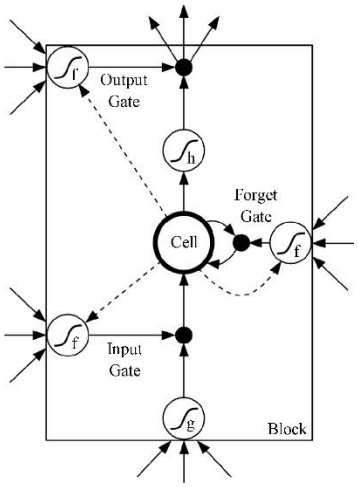

图 4.2 LSTM 隐含层单元结构[22][22]

4.4 预测结果评价指标

为了 SEIR 模型更好的拟合疫情数据曲线，我们将每阶段曲线分为 6 段，分别设置参数。在 ARIMA 模型和 LSTM 模型中，在每一期数据中，我们将前 80%数据划分为训练集，其余 20%数据划分为测试集。我们通过绘制真实值和预测值的曲线图来对比评估他们的预测效果。三类模型都属于定量预测模型，而且真实值和预测值都出现了取 0 的情况，因此我们选择了 RMSE(Root-mean-square error，均方根误差)、MAE（Mean Absolute Error，平均绝对误差）和 R^2 三个指标了对模型的预测效果进行评估。RMSE 是用来衡量真实值与预测值偏差的指数，值越小拟合效果越好，常被用作判断曲线的拟合程度的好坏。

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2}$$（7）

MAE 是预测值与真实值的误差绝对值的平均值，效果与 RMSE 相同，但它可以更直观的反应出预测曲线与真实曲线的差距。

$$MAE = \frac{1}{m}\sum_{i=1}^{m}|y_i - \hat{y}_i|$$（8）

R^2 是指拟合优度，其值在[0,1]之间，越接近 1 说明模型的拟合效果越好。

$$R^2 = 1 - \frac{ss_{res}}{ss_{tot}} \quad (9)$$

SS_{res} 表示真实值与预测值的差的平方和，SS_{tot} 表示数据的平方差。

4.5 Matlab 建模流程

本文使用 Matlab 2018a 软件进行建模，利用 Matlab 对三个模型分别进行建模。

建模步骤如下：

（1）导入数据

从 excel 表中导入某个国家某时期的新冠肺炎疫情数据。

（2）ARIMA 模型

对数据进行平稳性检验，若数据不平稳，需要对其进行差分处理，直至通过检验。通过 AIC 准则确定参数 p 和 q，计算预测值。

（3）LSTM 模型

将数据划分为训练集和测试集，对数据进行预处理，将原始数据转化为零均值和单位方差的数据，求解器设置为 adam，共 250 轮训练。指定初始学习率 0.01，在 125 轮训练后通过乘以因子 0.2 来降低学习率。计算预测值并去标准化。

（4）SEIR 模型

设定参数，将数据根据每个国家的具体情况划分为 6 个区段分别用模型进行拟合，计算得到累计确诊人数。

（5）绘图

绘制上述 3 个模型得到的预测值与真实值的曲线图。

（6）计算评价指标

分别计算各个模型 RMSE、MAE 和 R^2 的值。

第五章 疫情预测结果与讨论

5.1 初期预测结果

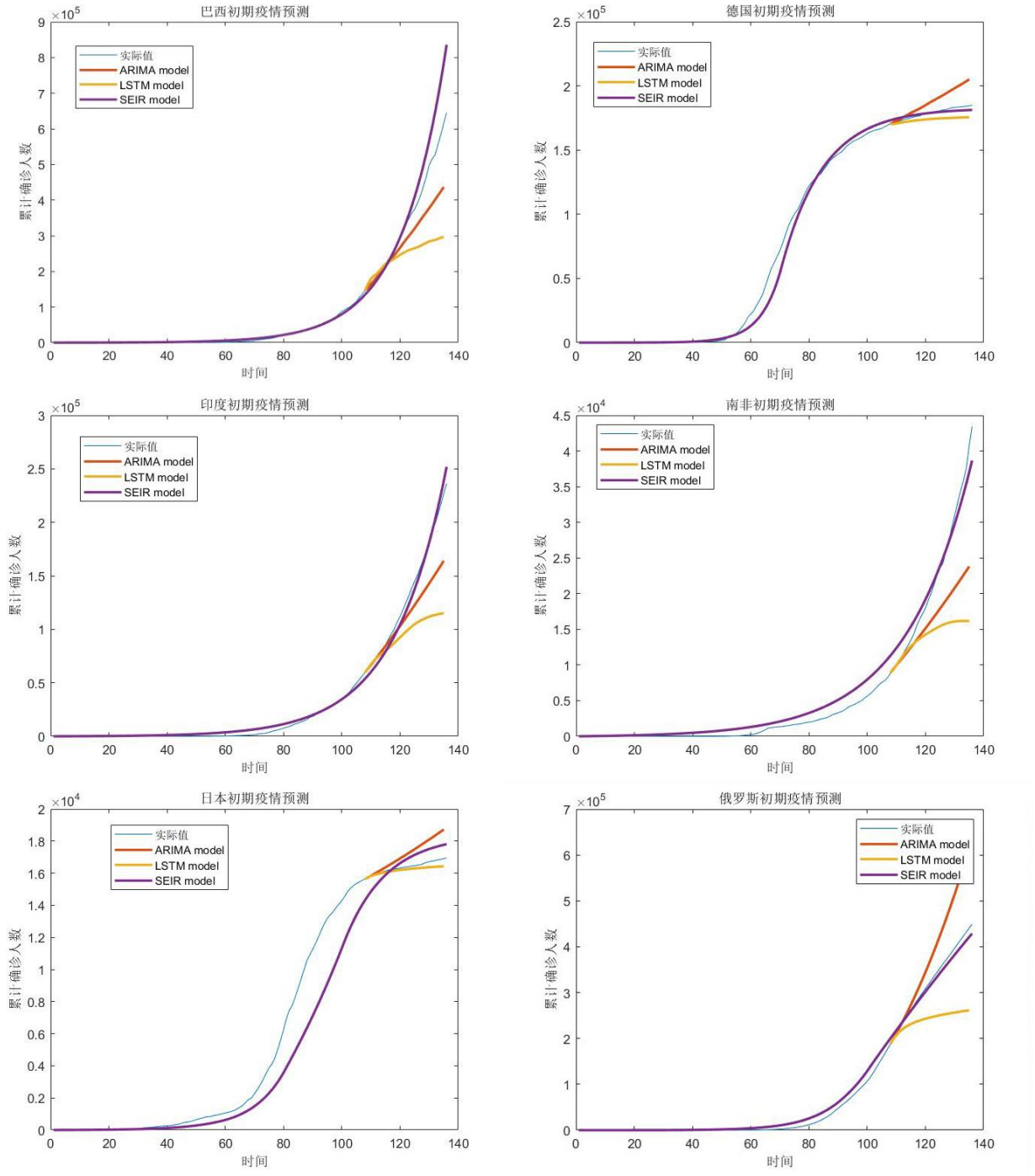

图 5.1 初期预测结果

初期的预测结果如上图，从图中可以看出，巴西、印度，南非和俄罗斯四个国家在疫

情爆发的前四个月中，累计确诊病例几乎呈现出指数式的上升，这也反应出当地政府可能对疫情没有采取有效措施，以致疫情在自然状态下扩散。LSTM 模型对这四个国家的预测较为保守，与实际情况出现了一定程度的偏差。ARIMA 模型的拟合效果也差强人意，仅能显示出其趋势，而与实际确诊人数还有差距。相较而言，SEIR 模型在这种情况下比较完美的拟合了疫情发展情况，这是因为 SEIR 模型本来就是描绘传染病自然传播情况的一类模型，这些国家的初期疫情缺乏人工干预手段，因此 SIER 模型可以很好的预测疫情发展趋势。

对德国和日本来说，这两个国家的疫情曲线都呈现出先急速上升又趋于平缓，这说明两个国家都采取了有效措施控制疫情，使得其没有进一步传播。在这种情况中，LSTM 模型的表现更为优秀，较为精确的表现出了后期疫情的发展趋势。ARIMA 模型对疫情发展呈现悲观态度，这可能是因为前期过快的增长影响了模型，ARIMA 模型中无法加入人工干预因素，只能依赖过去的数据。SEIR 模型在这种情况下，拟合的程度取决于是否可以获得精确的各项参数，这往往依赖于科研人员对病毒的了解程度、各项措施的实施力度、民众的防护意识等等，参数越精确，拟合程度就越好，但是一般情况下在疫情发展初期，我们很难获取这些信息。下表是三种模型初期疫情预测的各项指标。

表 5.1 疫情初期预测各项指标

国家	模型	RMSE	R^2	MAE
巴西	ARIMA	94553.405031	0.925720	71931.418389
	LSTM	128973.989162	0.861796	96306.638244
	SEIR	32831.538452	0.954890	11316.842168
日本	ARIMA	951.633228	0.996264	763.573130
	LSTM	286.471154	0.999661	216.760784
	SEIR	1638.528105	0.944212	1013.760494
俄罗斯	ARIMA	75861.230865	0.936843	56081.763908
	LSTM	90941.824741	0.909237	78199.206192
	SEIR	9800.249664	0.994691	6946.937682
德国	ARIMA	10491.558759	0.996347	8521.682975
	LSTM	2489.519026	0.999794	2318.350118
	SEIR	5732.024522	0.994508	3202.485793
南非	ARIMA	8546.995521	0.849027	6622.496735
	LSTM	11856.274882	0.709485	8878.860914
	SEIR	1347.965416	0.981085	1038.614910
印度	ARIMA	33715.928795	0.933193	26233.941124
	LSTM	61585.585487	0.777101	48583.908078

SEIR	4035.767118	0.995179	2881.052703

从表中可以看出，初期疫情预测中 SEIR 模型的表现最好，R^2 多次达到了 99%，说明拟合效果很好。LSTM 模型的表现最差，最低只有 70%。

5.2 中期预测结果

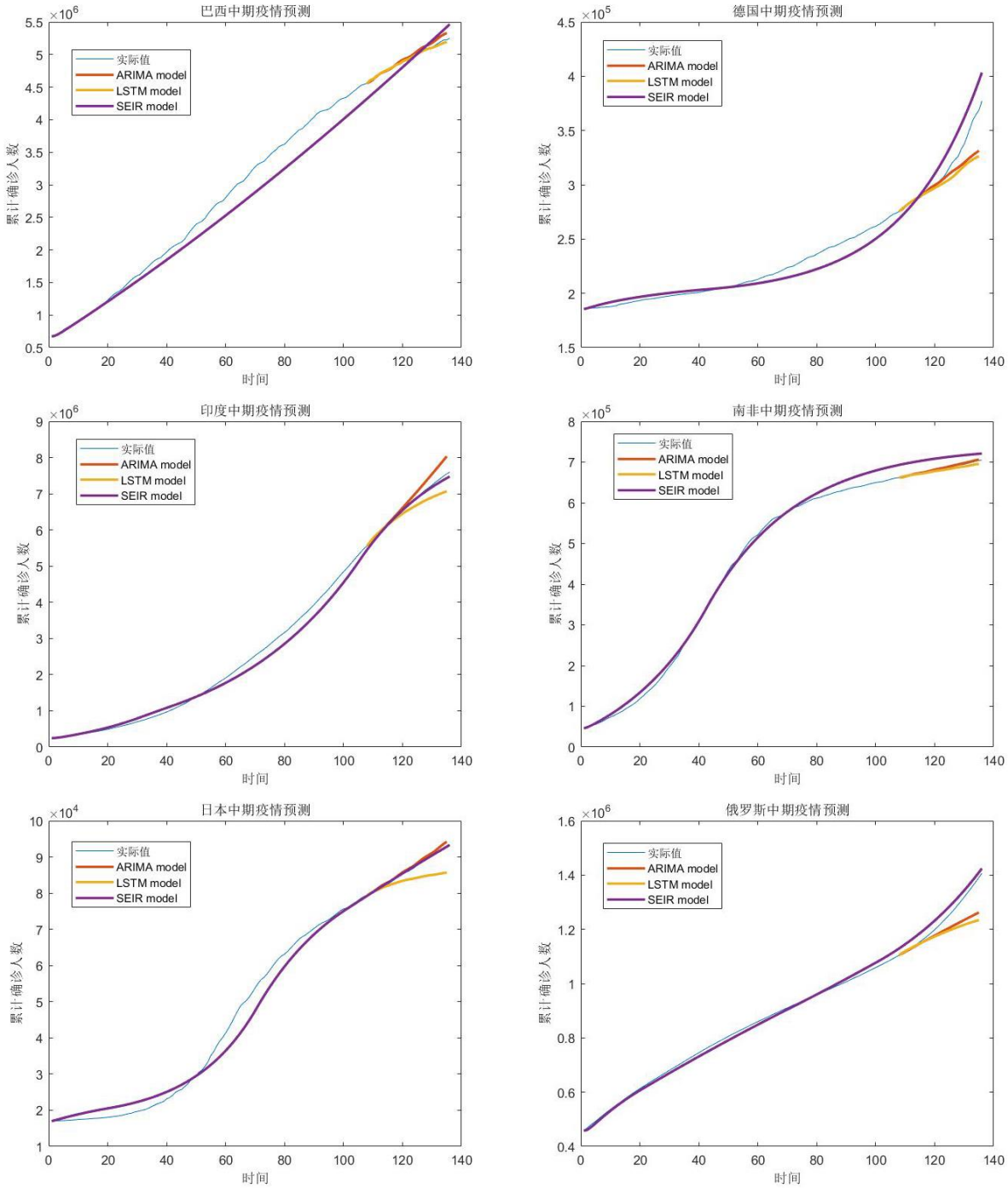

图 5.2 中期预测结果

中期预测结果如上图，初期疫情控制比较好的德国和日本在中期都出现了第二波疫情，不同的是日本仍然控制住了疫情，而德国国内疫情却出现指数上升趋势，有爆发迹象。印度、俄罗斯和巴西的确诊人数延续了初期的趋势，仍然呈现出上升态势，只是上升趋势有所缓和，不再呈现出指数式的上升。南非确诊病例在经历了近 200 天的增长后出现了缓和，国内疫情得到一定程度的控制，说明政府采取了有力措施。

从总体拟合情况来看，SEIR 模型在这一阶段的拟合效果并不好，主要原因是因为疫情经过一轮爆发后，各国纷纷采取应对措施，人为干预的因素大大增加，如果不能及时获取各项精确的参数，SEIR 模型很难反应出疫情走向。ARIMA 模型与 LSTM 模型在这一阶段的拟合情况都很好，两个模型相较而言这一阶段 ARIMA 模型的表现更好，在南非和巴西的预测中几乎精确的预测了病例增长情况；LSTM 模型的预测结果都有保守倾向，总是位于真实曲线的下方。

表 5.2 疫情中期预测各项指标

国家	模型	RMSE	R^2	MAE
巴西	ARIMA	34651.759593	0.999885	27258.289578
	LSTM	74838.700258	0.999463	66012.407992
	SEIR	266139.039366	0.965824	203325.960149
日本	ARIMA	293.961193	0.999977	228.916019
	LSTM	3240.765350	0.997173	2584.019723
	SEIR	2742.623037	0.989803	1985.359690
俄罗斯	ARIMA	69179.956466	0.983405	54319.916691
	LSTM	30349.339850	0.996806	20949.614565
	SEIR	16511.787985	0.995238	13914.464904
德国	ARIMA	18081.601000	0.971120	11804.176719
	LSTM	22396.138450	0.955693	15030.997003
	SEIR	10331.889185	0.952504	7877.485247
南非	ARIMA	1024.927789	0.999996	807.040204
	LSTM	12367.612957	0.999408	10063.127158
	SEIR	18082.088429	0.993625	14850.717320
印度	ARIMA	177715.079435	0.998877	133729.062856
	LSTM	357938.941508	0.989672	295080.362966
	SEIR	176660.922638	0.994412	132387.730184

从表中可以看出，中期 ARIMA 与 LSTM 模型的拟合效果都非常好，SEIR 模型虽然误差不大，但是较为依赖参数的精确性与及时性，在事后拟合的情况下可能会出现较好的效果，但是在疫情发生过程中可能难以很精确的反映其变化趋势。

5.3 后期预测结果

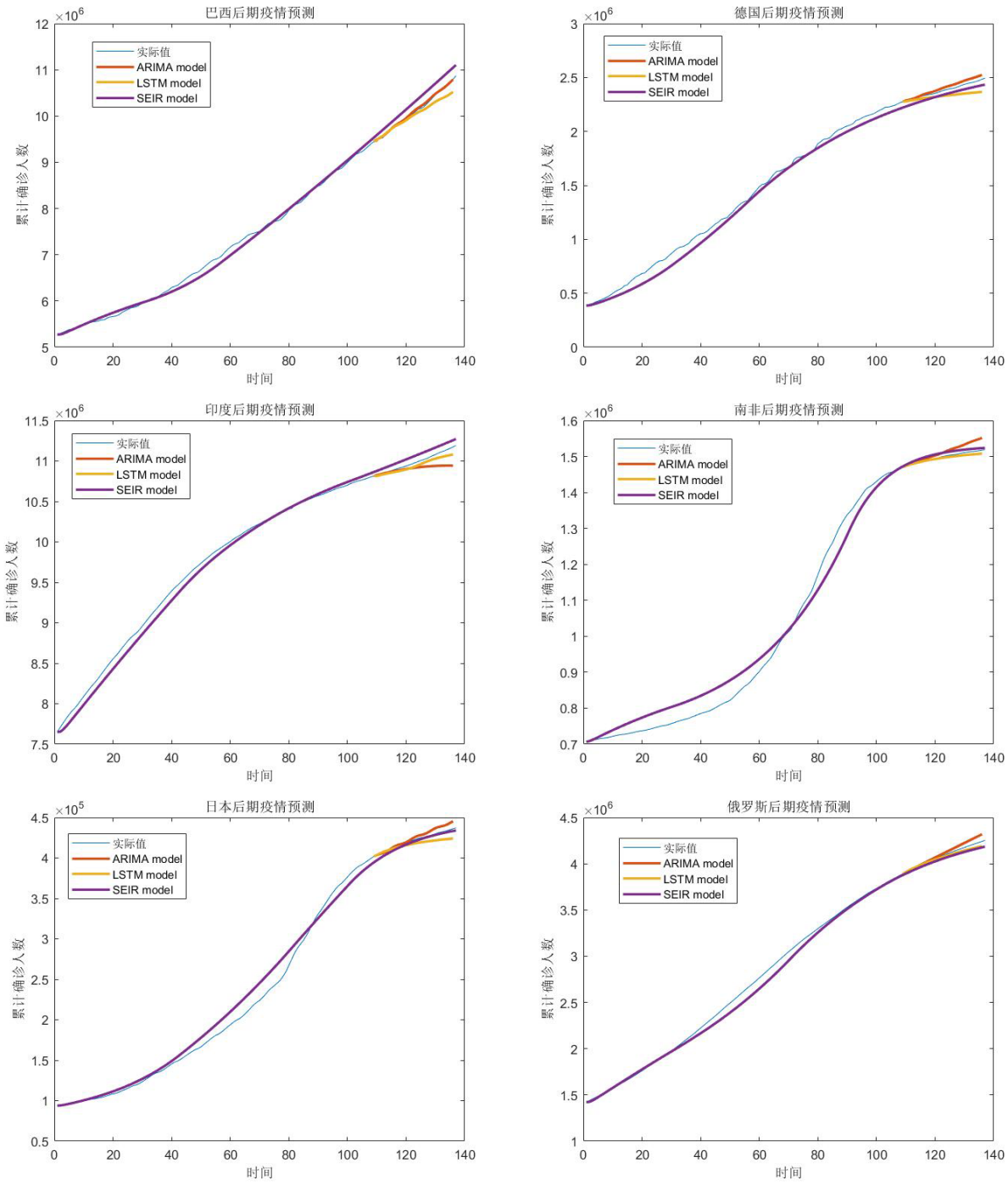

图 5.3 后期预测结果

从上图可以看出，上期疫情控制较好的两个国家：南非和日本，在后期仍然出现了疫情反弹，政府的控制措施使得其增长曲线再次趋于平缓，对于这类情况，从中期的预测结果我们可以得到，ARIMA 模型与 LSTM 模型对这类情况的拟合结果非常好，SEIR 模型的预测结果稍差一些。从结果来看，这个结论在后期同样适用。

巴西和俄罗斯的疫情仍然没有得到控制，确诊人数继续呈现上升趋势。对这类情况，三种模型的预测效果都很好，由于增速放缓，也没有出现初期 ARIMA 与 LSTM 模型偏差过大的情况。

印度和德国确诊人数的增长速度放缓，但疫情危机仍然没有解除，疫情没有得到有效控制，仍然呈现出上升趋势，可能的原因是政府措施不够强力。ARIMA 与 LSTM 模型对印度的预测结果有偏差，印度疫情在末期出现了反弹，可以看出模型对突发的增长无法处理，这也是 ARIMA 与 LSTM 模型很难进行长期预测的原因。

表 5.3 后期疫情预测各项指标

国家	模型	RMSE	R^2	MAE
巴西	ARIMA	33676.400043	0.999921	27717.397863
	LSTM	175926.835690	0.997836	141919.725663
	SEIR	126291.117419	0.994341	98124.019252
日本	ARIMA	3922.458806	0.999800	3129.153439
	LSTM	4762.515059	0.999705	4037.513402
	SEIR	10086.556805	0.993279	7571.794959
俄罗斯	ARIMA	29795.114053	0.999786	22804.000665
	LSTM	17546.164194	0.999926	11474.940463
	SEIR	56608.543929	0.996086	43118.332190
德国	ARIMA	20468.634948	0.999819	17459.617033
	LSTM	60025.785162	0.998444	49823.234867
	SEIR	58804.309591	0.992423	51745.284829
南非	ARIMA	16350.476076	0.999469	12455.992348
	LSTM	6087.199685	0.999926	5313.988110
	SEIR	33897.977895	0.988414	27757.355966
印度	ARIMA	119684.858575	0.997167	94982.176867
	LSTM	194147.991161	0.992546	168526.102596
	SEIR	76872.586943	0.994070	67179.566108

从表中可以看出本期预测效果都非常好，在疫情发展的中后期，LSTM 和 ARIMA 模型的预测能力大幅度提高，可以很好的预测短期疫情发展。SEIR 模型仍然存在参数获取的问题，在获取到精确参数的情况下，可以精确的反应模型。

第六章 结论和建议

研究发现：第一，SEIR、ARIMA和LSTM模型都可以实现效果非常好的疫情预测结果；第二，各类模型都有其适用场景，在合适的情景下使用正确的模型才能达到最好的效果。具体来说，三个模型都有各自的优缺点，对于SEIR模型来说，它最适合描述自然状态下的疫情发展态势，在引入其他变量后也可以很好的描述其他情况，但需要对传染病的特性、防控措施可能产生的效果以及国内人民的防护意识都要有清晰的了解，这样才能获取准确的各项参数，以达到预测疫情走向的目的，该模型更适合专业人员使用，而且在疫情初期很难获取所需要的各项参数，只能大概判断疫情走向。

对ARIMA模型来说，它可以在大多数情况下对疫情做出合理预测，是一类较为可靠的预测模型，而且不需要获取参数，仅需要时间和数据就可以很好的预估疫情发展，这类模型也被广泛的应用到传染病预测领域，其缺陷就是无法处理突发情况，当疫情出现被强力防控措施遏制或是出现快速增长的情况时，ARIMA模型很难准确地做出判断，这也是ARIMA模型只能做短期或中期预测的原因。

LSTM模型在疫情中后期的表现非常好，它是一种神经网络模型，因而它对数据的要求更低，几乎不需要做任何处理，也无需任何参数就可以做出效果非常好的预测。神经网络预测模型逐渐成为传染病预防的热点领域，它的适应性和学习性都很强，几乎可以适用于任何情况，但是我们在研究的过程中也发现了它的一些缺陷，同样无法准确预测快速增长的疫情趋势，在全部的预测结果中，LSTM模型的预测曲线总是位于真实曲线的下方，这也反应出该模型的预测结果比较保守，该模型也并不适用于长期预测，只能短期内预判疫情的走势。

从横向看，三个时期各个模型的预测准确率逐渐上升，在疫情爆发之初，关于病毒的各类信息尚不明确，其传播途径和感染性未知，感染人数呈现指数上升，这给模型预测带来了困难。由此可见早期的信息披露至关重要，及时准确的信息可以帮助我们对疫情的发展规模做出合理的判断。

综上，三类模型都有各自的适用场景，在信息较为匮乏的时候我们可以选择时间序列模型与神经网络模型，但只能做出短期预测；在信息充足的情况下我们可以选择SEIR模型，它不仅能描绘疫情走向，还可以比较不同防控措施的效果，为科学合理的制定疫情防控策略提供理论依据。

参考文献

[1] 余艳妮，聂绍发，廖青，等. 传染病预测及模型选择研究进展[J]. 公共卫生与预防医学，2018,29(05):89-92.
[2] 朱仁杰, 唐仕浩, 刘彤彤, 等. 基于改进SIR模型的新型冠状病毒肺炎疫情预测及防控对疫情发展的影响[J]. 陕西师范大学学报(自然科学版), 2020,48(03):33-38.
[3] Wenxiao J, Yi W, Yanpu L, et al. Integrating Multiple Data Sources and Learning Models to Predict Infectious Diseases in China.[J]. AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science, 2019,2019.
[4] 温亮, 黄清臻, 王志刚, 等. 运用ARIMA模型预测巴基斯坦新型冠状病毒肺炎疫情发展趋势的结果分析[J]. 解放军预防医学杂志, 2020,38(08):96-99.
[5] Jonas D, Johannes Z, Paul S F, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions.[J]. Science (New York, N.Y.), 2020,369(6500).
[6] 周涛, 刘权辉, 杨紫陌, 等. 新型冠状病毒肺炎基本再生数的初步预测[J]. 中国循证医学杂志, 2020,20(03):359-364.
[7] 杨雨琦, 孙琦, 王悦欣, 等. 重庆市新型冠状病毒肺炎(NCP)疫情分析与趋势预测[J]. 重庆师范大学学报(自然科学版), 2020,37(01):135-140.
[8] 梅文娟，刘震，朱静怡，等. 新冠肺炎疫情极限IR实时预测模型[J]. 电子科技大学学报，2020,49(03):362-368.
[9] 邢慧娴, 杨维中, 王汉章. 传染病预测[J]. 预防医学情报杂志, 2004(06):639-642.
[10] 王丙刚, 曲波, 郭海强, 等. 传染病预测的数学模型研究[J]. 中国卫生统计, 2007(05):536-540.
[11] 蔡洁, 贾浩源, 王珂. 基于SEIR模型对武汉市新型冠状病毒肺炎疫情发展趋势预测[J]. 山东医药, 2020,60(06):1-4.
[12] 范如国, 王奕博, 罗明, 等. 基于SEIR的新冠肺炎传播模型及拐点预测分析[J]. 电子科技大学学报, 2020,49(03):369-374.
[13] 王旭艳, 喻勇, 胡樱, 等. 基于指数平滑模型的湖北省新冠肺炎疫情预测分析[J]. 公共卫生与预防医学, 2020,31(01):1-4.
[14] 杨真真, 谢艳秋, 靳旭东, 等. 基于ARIMA时间序列模型的传染病发展趋势预测——以COVID-19为例[J]. 中国科技信息, 2021(Z1):70-72.
[15] 吴志强, 王波. 基于组合神经网络模型的新冠疫情传播预测分析[J]. 软件导刊, 2020,19(10):15-19.
[16] Anuradha T, Neeraj G. Prediction for the spread of COVID-19 in India and effectiveness of preventive measures.[J]. The Science of the total environment, 2020,728.
[17] 王志心, 刘治, 刘兆军. 基于机器学习的新型冠状病毒(COVID-19)疫情分析及预测[J]. 生物医学工程研究, 2020,39(01):1-5.
[18] 唐金芳, 王佑新, 农初师, 等. 2020年南宁市流行性腮腺炎ARIMA模型预测研究[J]. 实用预防医学, 2021,28(03):313-316.
[19] 彭月. ARIMA模型的介绍[J]. 电子世界, 2014(10):259.
[20] 王英伟，马树才. 基于ARIMA和LSTM混合模型的时间序列预测[J]. 计算机应用与软件，2021,38(02):291-298.

[21] 陈亮, 王震, 王刚. 深度学习框架下LSTM网络在短期电力负荷预测中的应用[J]. 电力信息与通信技术, 2017,15(05):8-11.
[22] Nbsp A, Graves. Supervised Sequence Labelling with Recurrent Neural Networks[M]. Springer, Berlin, Heidelberg.

致 谢

首先感谢曲宗希老师对我的悉心指导，从课题的选择到论文完成，老师都给予了我很大的帮助和支持，多次与我探讨了论文的结构和思路，并给我提出了许多可行性的意见和建议，全面细心的帮助我修改论文。

感谢管理学院全体老师，让我在前三年中收获了很多专业知识，你们严谨的治学精神和无私的奉献精神，不仅让我对知识有了更深层次的理解，而且还教会了我许多做人做事的道理。

感谢各位同学在我论文撰写过程中为我答疑解惑，谢谢他们的帮忙和鼓励。

最后感谢默默支持我的父母，他们的关心和理解是我前进的动力。

毕业论文（设计）成绩表

导师评语

论文选题符合专业培养目标，能够达到训练目标，选题有一定难度，具有一定创新型，工作量饱满，研究具有现实意义，文章篇幅完全符合学院规定，内容完整，层次清楚，有一定个人见解。文题相符，论文突出，撰扣主题。格式规范。

建议成绩 优秀　　指导教师（签字）

答辩委员会意见

答辩委员会负责人（签字）

成绩　　学院（盖章）

年　月　日

兰州大学本科生优秀毕业论文电子版使用授权书

《 突发传染病疫情预测模型对比分析研究——以新冠疫情为例 》是本人在兰州大学的本科毕业论文，现已通过答辩。本人作为此论文的著作权人，同意向兰州大学图书馆提交该论文的电子版和印刷本各一份。

根据《中华人民共和国著作权法》的规定，本人授权兰州大学图书馆对该论文电子版享有以下权力：（同意者画√）

1、同意提交全文。可以在

公开（半年）☑　　延时公开（1年□，2年□）

期限之后，由图书馆在校园网上提供全文浏览。

2、不同意提交电子版论文。□（选此项者，须由作者本人出具不能公开的证明，导师签字，院系所加盖公章。否则，电子版论文正常提交。）

图书馆承诺：

1、不对论文从事收集、保存、发布以外的其他活动；

2、未经著作权人同意，不得从事营利性活动。

院系：管理学院　　学号：320170914591

作者（授权人）签名：贾凯文　　时间：2021.5.25

被授权人：兰州大学图书馆