

修改说明

(1) 模型输入变量的选择与处理

论文仅使用了 HFRS 发病率历史数据作为模型输入,建议考虑将一些可能影响 HFRS 发病的气象、环境等外部因素纳入模型,以进一步提高预测精度。

对于输入变量的预处理,如缺失值填补、异常值处理等,文章缺乏详细说明,可以补充相关内容。

感谢专家意见。针对第一点意见“考虑将一些可能影响 HFRS 发病的气象、环境等外部因素纳入模型,以进一步提高预测精度”,我们深感认同其重要性及必要性。理想情况下的确应该纳入更多元化的外部变量优化模型,以期获得更精准的预测结果。然而,本文研究的是全国 HFRS 发病率预测,全国范围的气象、环境因素较为复杂,鉴于当前时间限制与资源约束,本文偏重在时间序列数据的处理、预测建模上,因此未纳入多种外部因素。尽管如此,我们已将这一方向视为未来研究的重点之一,这在文章“5 结语”的局限性及未来研究展望中重点提出,计划在后续工作中充分利用更广泛的数据源,以期构建更为综合、精细、高预测准确率的模型。

针对第二点“输入变量的预处理”相关意见,由于本文数据来源于作为国家人口健康科学数据共享平台数据中心之一的公共卫生科学数据中心,数据具体为 HFRS 逐月发病率资料,数据较为标准,没有缺失值与异常值,所做的数据预处理主要是数据资料的整合。文章进一步在“2.1 资料来源”处补充关于数据的说明与处理相关内容。

(2) 模型参数优化

论文中提到使用网格搜索等方法确定了某些模型参数,但对于其他参数的选择过程并未详细说明,可以补充优化过程的描述。

感谢专家意见。针对参数选择过程的详细说明,已在文章“3.3 STL-ARIMA 模型建立”、“3.5 NNAR 模型建立”、“3.6 LSTM 模型建立”处补充。其中,关于 LSTM 模型建立的参数选择,增加表 1 展示实验过程中,比较多个关键参数对预测准确率影响的结果,从而体现该模型构建中参数的优化过程。

1.文章一般以结语结束,指出存在的不足及下一步研究方向。目前与讨论相连,建议酌情微调。

感谢专家意见。文章重新将讨论中的段落切分为合适的部分,增加“5 结语”,指出研究局限性以及未来研究方向。

2.文章一般不超过 9000 字符数(目前 13000 余字符数),请适当控制一下篇幅。

感谢专家意见。文章正文已缩减至 4464 个字,6423 个字符,符合要求。

温馨提示:每篇论文需在文末加注利益冲突声明和作者贡献声明:“利益声明:所有作者均声明不存在利益冲突”“作者贡献:xxx 负责 xxxxx; xxx 负责 xxxxx; xxx 负责 xxxxx。”

文章已在文末加注利益冲突声明和作者贡献声明。

基于 SARIMA-LSTM 模型的肾综合征出血热发病率预测研究

唐诗诗¹ 李宇轩¹ 唐圣晟¹ 刘庆华² 周毅^{1*}

(¹中山大学中山医学院 广州 510080 ²中山大学附属第一医院肾内科 广州 510080)

【摘要】目的/意义 探究前沿技术在肾综合征出血热（HFRS）发病率预测中的应用，梳理、组合多种时序分析方法，并评价模型效果，筛选最佳模型。**方法/过程** 利用 2004-2020 年全国 HFRS 发病率数据，分析基于统计学方法的 SARIMA、STL-ARIMA、TBATS 模型，基于神经网络的 NNAR、LSTM 模型，基于三种加权方式的 SARIMA-LSTM 组合模型的预测效果，利用 RMSE、MAE、MAPE 综合评价模型效果。**结果/结论** SARIMA、LSTM 为单一模型中较优模型，RMSE、MAE 和 MAPE 分别为：0.01224、0.00981、18.43%，0.00998、0.00705、14.08%；SARIMA-LSTM 组合模型相较单一模型效果均有提升，基于误差倒数法的 SARIMA-LSTM 组合模型为最优模型，三项评价指标值分别为：0.00940、0.00519、9.32%，筛选出的最佳模型以及组合策略，有望为 HFRS 发病预警系统模型设计提供技术支持与参考。

【关键词】 肾综合征出血热；传染病监测预警；统计学模型；机器学习；SARIMA-LSTM 模型

Research on Hemorrhagic Fever with Renal Syndrome Incidence Prediction Based on the SARIMA-LSTM Model

TANG Shishi¹, LI Yuxuan¹, TANG Shengsheng¹, LIU Qinghua², ZHOU Yi^{1*}

¹ Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou 510080, China; ² Department of Nephrology, The First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China

【Abstract】 Purpose/Significance To investigate the application of cutting-edge technologies in predicting the incidence of Hemorrhagic Fever with Renal Syndrome (HFRS), to compile and integrate various time-series analysis methods, and to evaluate model performance in selecting the optimal model. **Method/Process** Utilizing national HFRS incidence data from 2004 to 2020, the predictive effectiveness of models based on statistical methods: SARIMA, STL-ARIMA, and TBATS, neural network approaches: NNAR, LSTM, and combined models of SARIMA-LSTM with three different weighting schemes were analyzed. The performance of these models is comprehensively assessed using RMSE, MAE, and MAPE. **Result/Conclusion** The SARIMA and LSTM models are identified as the superior individual models, with their respective performance

【作者简介】 唐诗诗，硕士研究生，发表论文 1 篇；通信作者：周毅，博士，教授，博士生导师。**【基金项目】** 国家重点研发计划项目（编号 2022YFC3601600），广东省自然科学基金项目（编号 2024A1515011989），广州市科技计划项目（编号 202206010028）。

metrics—RMSE, MAE, and MAPE—recorded as follows: 0.01224, 0.00981, and 18.43% for SARIMA; 0.00998, 0.00705, and 14.08% for LSTM. The combined SARIMA-LSTM model demonstrates enhanced performance compared to individual models. The SARIMA-LSTM model optimized using the reciprocal of error method is deemed the optimal model, achieving significantly reduced error measures with values of 0.00940 for RMSE, 0.00519 for MAE, and 9.32% for MAPE. The selection of this optimal model and the strategic combination approach bodes well to offer technical support and guidance for the development of an early warning system model tailored to forecasting HFRS outbreaks.

[Keywords] hemorrhagic fever with renal syndrome; infectious disease surveillance and early warning; statistical model; machine learning; SARIMA-LSTM model

1 引言

传染病暴发构成重大公共安全挑战，对人类社会造成深远影响。随着信息技术的发展，以数字化和网络化为标志的现代信息技术成为突发公共卫生事件现代化与智能化的基础。Error: Reference source not found。传染病预测模型通过分析疾病特性与传播模式，基于历史数据预测未来趋势，为防控措施提供理论依据，以信息化的方式推进应急管理的现代化^[2]。肾综合征出血热（Hemorrhagic Fever with Renal Syndrome, HFRS）是严重影响我国公共卫生的乙类传染病，近年来全国所有省市均有病例报告，发生范围不断扩大^[3]。全国累计报告病例达到 224 396 例，死亡病例 2068 例，年均发病率仍高达 0.93/10 万，这表明 HFRS 对全社会持续构成威胁，并且该疾病呈现复杂的地区分布特征，也加剧了疫情防控的复杂性^[4]。因此，开展更为精准的 HFRS 发病率预警模型研究显得尤为迫切与重要。

国内 HFRS 预测模型研究多针对局域地区，并且算法模型创新和疾病种类相对较少，多数研究仅使用 2~3 种模型预测方法^[5]，尤其在多模型比较、机器学习组合模型构建上的探索相对有限，难以满足高效、精确预测的需求。本研究创新地将 SARIMA-LSTM 组合模型应用到 HFRS 发病率预测中，为 HFRS 防控提供强有力的支持，并为传染病预警系统的优化及公共卫生政策制定提供科学依据。

2 材料与amp;方法

2.1 资料来源

本研究资料来源于公共卫生科学数据中心（<http://www.phsciencedata.cn/>），包括全国 HFRS 按地区分布的发病数、死亡数、发病率和死亡率数据，从平台按月整合 2004-2020 年全国 HFRS 逐月发病率资料。以 2004 年 1 月-2018 年 12 月全国 HFRS 发病率数据作为训练数据，以 2019 年 1 月-2020 年 12 月数据作为测试数据。选择均方根误差（Root Mean Square Error, RMSE）、平均绝对误差（Mean Absolute Error, MAE）、平均绝对百分比误差（Mean Absolute Percentage Error, MAPE）三项指标评估模型预测效果。

2.2 方法

本研究主要工具为 R 语言和 Python，实现了

SARIMA、STL-ARIMA、TBATS、NNAR 和 LSTM 模型的建立，基于预测效果组合建立了 SARIMA-LSTM 模型，最后综合三种损失函数评价模型效果。研究技术路线框架如图 1:

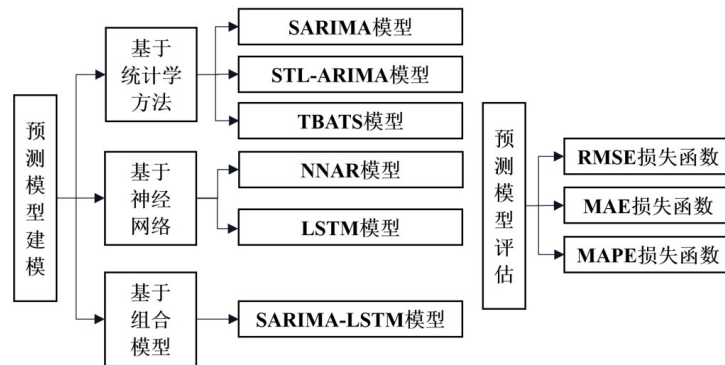


图 1 研究技术路线框架

2.2.1 SARIMA 模型

季节性自回归积分滑动平均 (Seasonal Auto-regressive Integrated Moving Average, SARIMA) 模型一般形式为 $SARIMA(p, d, q)(P, D, Q)_s$ 。本研究建模思路为: ①原始序列平稳性检验, 非平稳则进行差分运算; ②若原始序列或差分后序列平稳, 则展开白噪声检验; ③模型为非白噪声序列并且具有季节性, 则考虑拟合 SARIMA 模型, 进行模型定阶、参数估计和择优; ④拟合模型残差的白噪声检验; ⑤选择最佳模型展开 HFRS 发病率预测。

2.2.2 STL-ARIMA 模型

针对 HFRS 发病率的预测, 还有一种在时间序列预测领域中常见的思路, 即“先分解, 后预测”。STL (Seasonal-Trend decomposition procedure based on Loess) 是一种通用且稳健的时间序列分解方法^[6]。该方法将时序数据分解为季节、趋势和残差三个分量, 分别展开分析和预测。本研究利用 STL 分解数据, 利用 ARIMA 模型进行建模预测, 最终将三个分解项的预测结果加总为最终预测结果。

2.2.3 TBATS 模型

针对 HFRS 发病率时间序列的预测, TBATS 模型 (Trigonometric seasonality, Box-Cox transformation, ARMA errors, Trend and Seasonal components) 包含了傅里叶三角函数季节性、Box-Cox 转换、ARMA 残差、趋势项和季节项。相比 ETS 和 ARIMA 模型, TBATS 具有处理多种复杂季节性模式时间序列的优势^{Error: Reference source not found}, 其主要目标是用指数平滑对复杂季节因素建模, 能够对季节性因素进行不同程度的均衡^[8]。

2.2.4 NNAR 模型

神经网络自回归 (Neural Network Autoregression, NNAR) 模型可以被视为描述复杂非线性关系和函数形式的神经元或节点网络^[9]。该方法基于神经网络框架, 主要原理是利用过去观测值为输入, 预测未来观测值, 通常由多个前馈神经网络层组成, 以是否包含隐藏层区分为简单神经网络 (等价线性回归) 和含隐藏层的神经网络, 通过添加隐藏层, 神经网络则变得非线性。

2.2.5 LSTM 模型

近年来, 循环神经网络 (RNN) 常用于时间序列的分析与预测, 能够获得很高的效率^{Error: Reference source not found}。1997 年, 长短时间记忆 (Long Short Term Memory Network,

bmr.202407.00046V1

LSTM) 模型首次被提出^{Error: Reference source not found}，该模型主要解决梯度消失问题。LSTM 是基于门控制单元结构的深度学习模型，可以用于多种时序预测问题，本研究中，利用 LSTM 对 HFERS 发病率数据开展单变量时间序列预测。

2.2.6 SARIMA-LSTM 模型

SARIMA 模型可以有效捕捉序列长、短期趋势和周期特征，LSTM 模型则适用于处理数据集中复杂的非线性和动态特征，因此利用 SARIMA 和 LSTM 的组合模型预测 HFERS 发病率，可以同时处理线性和非线性的分量，从而有效提升模型预测能力。本研究选择并联模式组合模型，实现流程如下图 2 所示：

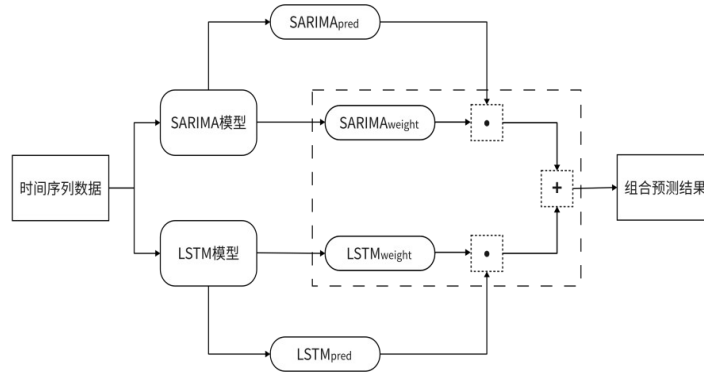


图 2 SARIMA-LSTM 并联组合模型流程

上述流程中， $SARIMA_{pred}$ 和 $LSTM_{pred}$ 分别表示 SARIMA 和 LSTM 模型的预测值； $SARIMA_{weight}$ 和 $LSTM_{weight}$ 分别表示模型组合时对应的权重。本研究采用的并联组合模型方法为定结构组合预测方法，其中单一预测模型的种类以及用于模型组合的权重均不变，通用表达式为：

$$Hybrid_{pred}[t] = \sum_{i=1}^n w_i f_i(t), i=1,2,\dots,n$$

(AUTONUM)

式 (1) 中， w_i 为第 i 个预测模型的权重 (其中， $\sum_{i=1}^n w_i=1$)， $f_i(t)$ 为第 i 个预测模型在 t 时刻的预测值， n 为参与组合的模型个数， $Hybrid_{pred}[t]$ 为加权不同模型得到的组合模型在 t 时刻的预测值。

对于模型权重的确定方法有以下三种，均满足 n 为参与组合的模型个数， w_i 为第 i 个预测模型的权重，且 $\sum_{i=1}^n w_i=1$ 的条件^{Error: Reference source not found}：

① 等权平均法：

该方法各预测模型的权重相等：

$$w_i = \frac{1}{n}, i=1,2,\dots,n$$

(AUTONUM)

② 误差倒数法：

该方法利用各预测模型的相对误差的倒数确定权重：

$$w_i = \frac{\varepsilon_i^{-1}}{\sum_{j=1}^n \varepsilon_j^{-1}}, i=1,2,\dots,n$$

(AUTONUM)

③Critic 客观权重赋值法：

该方法是一种基于训练数据的集成学习方法，根据预测结果与真实标签间的误差大小，分配组合权重：

$$w_i = \frac{RMSE_i^{-1}}{\sum_{j=1}^n RMSE_j^{-1}}, i=1,2,\dots,n$$

(AUTONUM)

3 结果

3.1 HFRS 发病率概况分析

本研究使用加法模型对原始数据进行时间序列分解，序列分解图如下图 3。由季节项分解得知 HFRS 呈明显的季节性，并且以 12 个月为周期，这与大多数乙类传染病的规律是相似的，并且每年有两个高峰期，第二个高峰的发病率明显高于第一个，也即 HFRS 常见发病于冬春季节。

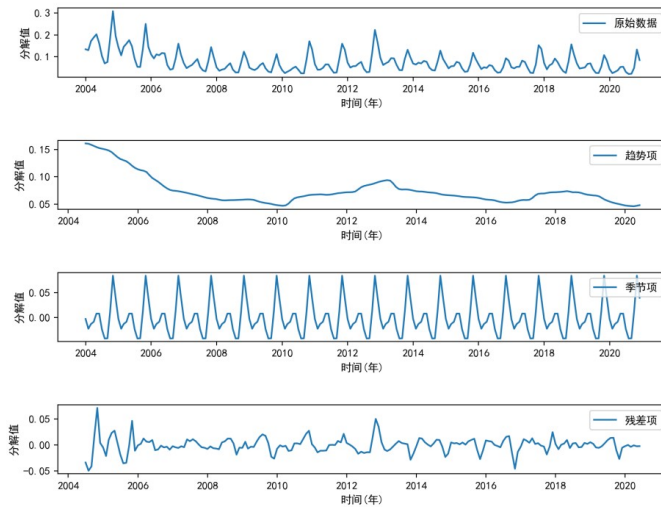


图 3 2004-2020 年全国 HFRS 发病率时间序列分解图

3.2 SARIMA 模型建立

本研究利用 ADF 和 KPSS 两种统计方法来检验序列平稳性，结合两种方法得知序列非平稳。由 Box-Ljung Q 检验得到序列为非白噪声，考虑其季节性，对数据进行以 12 个月为周期的季节差分。根据序列特征和滞后项，并结合 SARIMA 模型相关参数的取值一般不超过 5 的规律^[13]，本研究利用 R 中 auto.arima() 函数，在默认区间内自动搜索参数，设定搜索的 p、q、P、Q 的最大值为 5，设定 stepwise=TRUE 展开逐步法选择模型，搜索区间内的每一对由 KPSS 检验确认的 d 和 D 都将使用 ARIMA(p, d, q)(P, D, Q)s 模型拟合给

bmr.202407.00046V1

定的时间序列数据，并使用信息准则确定最佳模型。

得到 HFRS 发病率数据的最优模型为 SARIMA(2,1,1)(1,1,0)₁₂。其中最佳模型对应的 AIC=-1017.626，Box-Ljung Q 检验的 p-value=0.906，残差为白噪声。得到拟合模型诊断图见图 4。由该图可知，模型拟合得到的序列残差随着时间的变化基本稳定；KDE 线走势紧随标准正态分布，正态 QQ 图也印证了残差是正态分布的；最后由残差序列自相关图显示其不存在自相关。该模型具有较好的拟合效果，可用于进一步预测。

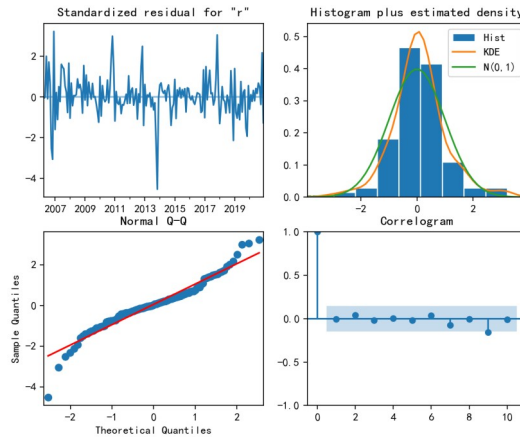


图 4 SARIMA 模型诊断图

真实值与预测值的比较见图 5。

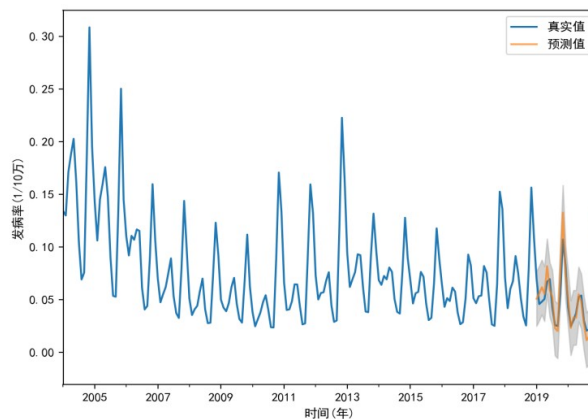


图 5 测试集上 SARIMA 模型真实值与预测值比较

3.3 STL-ARIMA 模型建立

本文使用 R 中 stlf() 函数实现 STL-ARIMA，利用网格搜索法确认 s.window 的最优取值为 5，自动选参函数在默认区间搜索得到最佳模型为 ARIMA(2, 1, 1)，AIC=-1230.45，残差序列 Box-Ljung Q 检验的 p-value=0.773，为白噪声。真实值与预测值的比较见图 6。

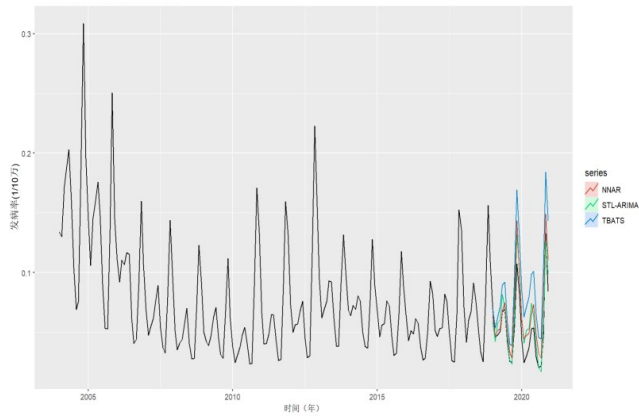


图 6 测试集上 STL-ARIMA、TBATS、NNAR 模型真实值与预测值比较

3.4 TBATS 模型建立

本研究利用 R 中 `tbats()` 函数建立 TBATS 模型，根据 AIC 最小化原则自动确定参数得到：TBATS(0.295, {3,0}, 1, {<12,5>})，其中 0.295 为 Box-Cox 变换的参数 λ ，ARMA 模型的 $p=3$ 、 $q=0$ ，季节周期长度为 12，傅里叶级数为 5，模型对应的 AIC=-737.60，残差序列 Box-Ljung Q 检验的 $p\text{-value}=0.884$ ，为白噪声。真实值与预测值的比较见图 6。

3.5 NNAR 模型建立

对于 NNAR 模型，由于 HFRS 发病率数据具有季节性，因此增加同一个季节的最后观测值为输入。利用 R 中 `nnetar()` 函数建模，通过交叉验证法得到 λ 为 0.6，函数拟合得到最佳模型为 NNAR(4,1,15)[12]。残差序列 Box-Ljung Q 检验的 $p\text{-value}=0.155$ ，为白噪声。真实值与预测值的比较见图 6。

3.6 LSTM 模型建立

本研究利用 Python 构建 LSTM 模型，建模步骤主要为：①数据准备：将数据以 12 个为一组切分，形成多个时间窗口，将这些时间窗口作为输入数据，下一个时间点的的数据则作为输出，正则化处理并划分训练集、测试集；②模型构建：形成[样本，时间步长，特征数]格式的数据，为减少计算复杂度设置时间步长为 1，特征为周期 12；③模型训练：表 1 显示了关键参数对训练集 RMSE 的影响，通过比较确定最佳参数，其中，隐藏层中设置 6 个神经元、批量大小设置为 1、设置优化器为 Adam，模型的损失函数为均方误差；④模型预测：利用已训练模型对测试数据展开预测；⑤误差计算：将预测数据反归一化，即转换为同一单位，计算相应误差指标。

表 1 关键参数对预测准确率的影响

隐藏层神经元	批量大小	优化器	RMSE
5	1	Adam	0.01064
6			0.01030
7			0.01116
6	1	Adam	0.01030
	16		0.01579
	32		0.01615
	64		0.02319
6	1	SGD	0.01755
		Adam	0.01030

bmr.202407.00046V1

Adagrad 0.02968
RMSprop 0.01109

真实值与预测值的比较见图 7。

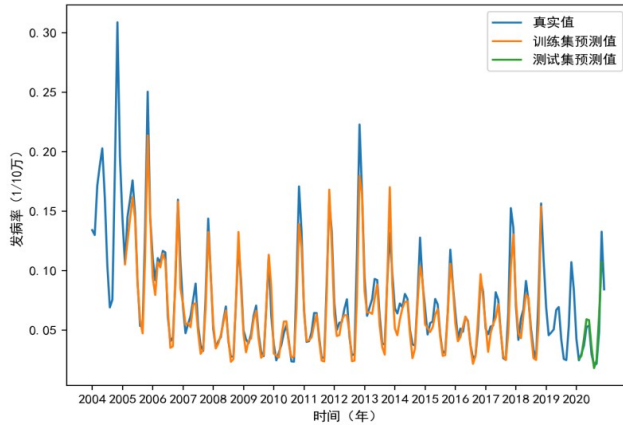


图 7 测试集上 LSTM 模型真实值与预测值比较

3.7 SARIMA-LSTM 模型建立

本研究利用 Python 构建组合模型，基于等权平均法、误差倒数法和 Critic 客观权重赋值法三种组合方法的模型分别命名为并联组合 1、并联组合 2 和并联组合 3，三种并联组合模型的权重见表 2。

表 2 三种并联组合模型权重

	并联组合 1		并联组合 2		并联组合 3	
模型	SARIMA	LSTM	SARIMA	LSTM	SARIMA	LSTM
权重	0.5	0.5	0.47	0.53	0.45	0.55

单一模型以及 SARIMA-LSTM 组合模型的预测结果见图 8。

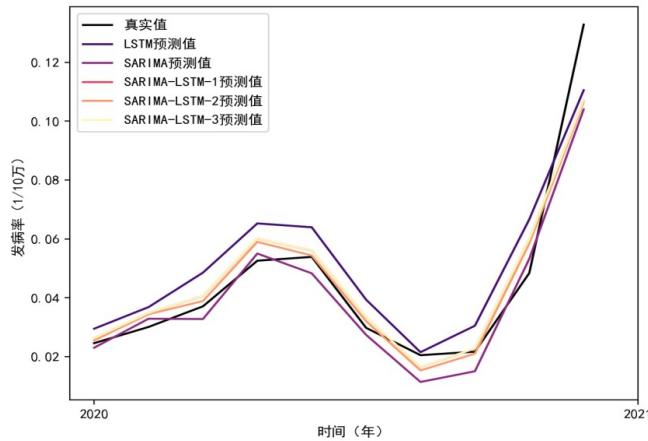


图 8 测试集上 SARIMA-LSTM 组合模型真实值与预测值比较

3.8 预测模型比较与评价

分别计算各个模型在测试集上的 RMSE、MAE 和 MAPE 三项指标值，评价各模型预测效果，得到不同模型的评价指标见下表 3。比较 HFRS 预测模型效果时，通常 $RMSE < 0.01$ 认为该模型的预测效果较优，RMSE 越小，预测效果越好；Pao HT^{Error: Reference source not found}提出 MAPE 介于 5%~10%时预测模型高度准确，介于 10%~20%的模型认为较好，在 20%~50%时表示模型可接受；三项指标均满足：值越小，预测效果越好。

bmr.202407.00046V1

通过以上比较规则，综合图像以及表 3 中的指标值可知，对于单一模型，STL-ARIMA、TBATS、NNAR 模型的效果较为中等；SARIMA 和 LSTM 模型在 HFRS 发病率预测上的各项指标都较优，其中 LSTM 已经满足较优模型的指标范围。对于 SARIMA-LSTM 并联组合模型，三种组合模型预测的各项指标相较单一模型都有一定的提升，其中达到最佳效果的是并联组合 2，即以误差倒数法确定权重的 SARIMA-LSTM 并联组合模型。

表 3 预测模型在测试集上的预测效果评价

模型	评价指标		
	RMSE	MAE	MAPE (%)
SARIMA	0.01224	0.00981	18.43
STL-ARIMA	0.01289	0.01069	22.69
TBATS	0.03720	0.03314	37.29
NNAR	0.01660	0.01357	30.27
LSTM	0.00998	0.00705	14.08
并联组合 1	0.00941	0.00522	9.55
并联组合 2	0.00940	0.00519	9.32
并联组合 3	0.00942	0.00520	9.43

4 讨论

中国在国际公共卫生界占据了重要作用，而传染病是中国的主要病因之一^{Error: Reference source not found}，紧紧围绕多渠道全维度传染病风险自动识别与早期智能探测预警^[16]，加强传染病监测数据的时序预测模型研究，对于实时监控疫情动态并迅速实施防控举措至关重要。本研究聚焦于探寻 HFRS 发病率预测的最佳模型，比较基于传统统计学方法的 SARIMA、STL-ARIMA 和 TBATS 模型，基于神经网络的 NNAR、LSTM 模型，和 SARIMA-LSTM 组合模型在测试集上的预测效果。

实验结果显示，单一模型中 LSTM 和 SARIMA 模型的三项指标值均较低，相较已有研究的实验结果^[17,18]，本研究由于训练数据增加以及多种参数优化操作，使得预测误差更低，均在 0.01 附近；三种方法得到的组合模型相较单一模型预测效果均有提升，RMSE 均小于 0.01，MAPE 均小于 10%，属于高度准确的预测模型，最佳组合模型为基于误差倒数法的 SARIMA-LSTM 模型。可以得出结论，SARIMA 与 LSTM 都是较好的 HFRS 发病率预测模型，并且两者的组合能够在单一模型基础上进一步降低预测误差。

5 结语

通过建立 SARIMA-LSTM 模型，可以同时利用 SARIMA 对周期、趋势的敏感性以及 LSTM 的非线性序列预测能力，实现预测精度的提升。SARIMA-LSTM 组合模型以及这种建模思路可以拓展到更多传染病的发病预测上。本研究仍然存在一定局限性：①数据受限于公开资源，可能一定程度上影响样本的充分性和时效性；②优化的串联组合、非线性组合^[19]、集成学习^[20]等组合建模技术有待进一步探索；③本研究未深入分析地域性流行异质性及气象、环境等外部因素对 HFRS 动态的影响。

未来，将进一步拓宽数据来源，并考虑多维度数据，纳入多种外部因素，以优化

bmr.202407.00046V1

HFRS 预测模型。从而进一步强化疫情的早期识别与反应能力，为医疗卫生资源配置的科学决策提供坚实支撑，从根本上助力传染病的防控与大流行的阻断。

利益声明：所有作者均声明不存在利益冲突。

作者贡献：唐诗诗负责实验设计、数据收集、数据分析、论文写作；李宇轩负责数据分析协助、论文修改；唐圣晟负责论文修改；刘庆华负责专业指导与论文审阅；周毅负责论文指导与审阅。

参考文献

- [1] Shapiro J S, Mostashari F, Hripcsak G, et al. Using health information exchange to improve public health[J]. *American journal of public health*, 2011, 101(4): 616-623.
- [2] Xiong L, Hu P, Wang H. Establishment of epidemic early warning index system and optimization of infectious disease model: Analysis on monitoring data of public health emergencies[J]. *International Journal of Disaster Risk Reduction*, 2021, 65: 102547.
- [3] Hsieh C J, Li C W, Cheng C A, et al. Epidemiologic characteristics of domestic patients with hemorrhagic fever with renal syndrome in Taiwan: A 19-year retrospective study[J]. *International Journal of Environmental Research and Public Health*, 2020, 17(15): 5291.
- [4] 邓雪飞,杜珊珊,黄晓霞,等.2004-2021年中国肾综合征出血热报告病例流行病学特征分析[J].*疾病监测*,2023,38(01):70-74.
- [5] 何琪乐,张瑾瑶,吴卓存,等.基于互联网数据的传染病预测模型研究进展[J].*医学信息学杂志*,2024,45(02):32-37.
- [6] Cleveland R B, Cleveland W S, McRae J E, et al. STL: A seasonal-trend decomposition[J]. *J. Off. Stat*, 1990, 6(1): 3-73.
- [7] Xiao Y, Li Y, Li Y, et al. Estimating the long-term epidemiological trends and seasonality of hemorrhagic fever with renal syndrome in China[J]. *Infection and drug resistance*, 2021: 3849-3862.
- [8] Zhao D, Zhang H. The research on TBATS and ELM models for prediction of human brucellosis cases in mainland China: a time series study[J]. *BMC Infectious Diseases*, 2022, 22(1): 934.
- [9] Perone G. Comparison of ARIMA, ETS, NNAR, TBATS and hybrid models to forecast the second wave of COVID-19 hospitalizations in Italy[J]. *The European Journal of Health Economics*, 2021: 1-24.
- [10] Torres J F, Martínez-Álvarez F, Troncoso A. A deep LSTM network for the Spanish electricity consumption forecasting[J]. *Neural Computing and Applications*, 2022, 34(13): 10533-10545.
- [11] Hochreiter S, Schmidhuber J. Long short-term memory[J]. *Neural computation*, 1997, 9(8): 1735-1780.
- [12] 杨艳,黄晴,龙思,等.基于 ARIMA-LSTM 的货运量组合预测方法研究[J].*交通科学与工程*, 2022, 38(02): 102-108.
- [13] Hyndman R J, Khandakar Y. Automatic time series forecasting: the forecast package for R[J]. *Journal of statistical software*, 2008, 27: 1-22.
- [14] Pao H T. Forecasting energy consumption in Taiwan using hybrid nonlinear models[J]. *Energy*, 2009, 34(10): 1438-1446.
- [15] Jia W, Wan Y, Li Y, et al. Integrating multiple data sources and learning models to predict infectious diseases in China[J]. *AMIA Summits on Translational Science Proceedings*, 2019, 2019: 680.
- [16] 赵嘉,苏雪梅,赵自雄,等.新形势下疾病预防控制信息化需求及应用场景分析[J].*医学信息学杂志*,2024,45(02):65-70.
- [17] 刘天,姚梦雷,侯清波,等.7种时间序列模型对全国肾综合征出血热发病率预测效果比较[J].*中国媒介生物学及控制杂志*,2022,33(04):548-554.
- [18] 刘天,姚梦雷,侯清波,等. SARIMA、GAM 和 LSTM 在肾综合征出血热预测中的应用效果比较[J].*中华地方病学杂志*, 2022, 41(9): 709-714.
- [19] Ruiz-Aguilar J J, Turias I J, Jiménez-Come M J. Hybrid approaches based on SARIMA and artificial neural networks for inspection time series forecasting[J]. *Transportation Research Part E: Logistics and Transportation Review*, 2014, 67: 1-13.
- [20] 王荣政,廖贤艺,陈湘萍,等.基于集成学习融合模型的血糖预测[J].*医学信息学杂志*,2019,40(01):59-62+84.