

时间序列广义回归神经网络模型 在病毒性肝炎发病率预测中的应用

孙亚军^{1,3}, 刘天², 姚远³

1. 珠海市第三人民医院, 广东 珠海 519000; 2. 荆州市疾病预防控制中心;

3. 重庆市九龙坡区疾病预防控制中心, 重庆 400050

摘要:目的 本研究旨在介绍基于时间序列广义回归神经网络(generalized regression neural network, GRNN)模型在中国病毒性肝炎发病率预测中的应用, 评估其拟合及预测精度。方法 收集全国 2004—2019 年病毒性肝炎逐月发病率数据构建时间序列, 以 2004 年 1 月—2019 年 6 月数据为训练数据, 2019 年 7—12 月数据为测试数据, 分别建立 GRNN 和 SARIMA 模型, 预测 2019 年 7—12 月发病率并与测试数据比较。采用平均绝对百分比误差(MAPE)评价模型拟合及预测效果。结果 GRNN 模型对各类肝炎的拟合 MAPE 在 1.67%~21.22%之间, 预测 MAPE 在 2.26%~17.17%之间。与之相比, SARIMA 模型对各类肝炎的拟合 MAPE 在 3.84%~7.87%之间, 预测 MAPE 在 2.54%~48.89%之间, 其中甲肝的预测 MAPE 为 48.89%, 显示出较大的预测误差。结论 GRNN 模型在我国病毒性肝炎月发病率预测中表现优于 SARIMA 模型, 适合推广应用。

关键词: GRNN; SARIMA; 预测; 病毒性肝炎; 时间序列

中图分类号: R512.6 文献标志码: A 文章编号: 1003-8507(2024)23-4260-06

DOI: 10.20043/j.cnki.MPM.202406385

Application of time series generalized regression neural network model in predicting the incidence of viral hepatitis

SUN Ya-jun*, LIU Tian, YAO Yuan

*The Third People's Hospital of Zhuhai, Zhuhai, Guangdong 519000, China

Abstract: Objective To introduce the application of the time series generalized regression neural network (GRNN) model in predicting the incidence of viral hepatitis in China and to evaluate its fitting and predictive accuracy. **Methods** Monthly incidence data of viral hepatitis from 2004 to 2019 were collected to construct time series. Data from January 2004 to June 2019 were used as training data, while data from July to December 2019 served as testing data. Both GRNN and SARIMA models were established to predict the incidence from July to December 2019, and the predictions were compared with the testing data. The mean absolute percentage error (MAPE) was employed to assess the model's fitting and predictive performance. **Results** The fitting MAPE for the GRNN model across various types of hepatitis ranged from 1.67% to 21.22%, while the predictive MAPE ranged from 2.26% to 17.17%. In comparison, the SARIMA model's fitting MAPE for various types of hepatitis ranged from 3.84% to 7.87%, with a predictive MAPE ranging from 2.54% to 48.89%. Notably, the predictive MAPE for hepatitis A was 48.89%, indicating a significant prediction error. **Conclusion** The GRNN model outperformed the SARIMA model in predicting the monthly incidence of viral hepatitis in China, suggesting its suitability for broader application.

Keywords: GRNN; SARIMA; Prediction; Viral hepatitis; Time series

病毒性肝炎(viral hepatitis)是我国法定报告乙类传染病的一组疾病的统称,包括甲型病毒性肝炎(简称甲肝)、乙型病毒性肝炎(简称乙肝)、丙型病毒性肝炎(简称丙肝)、戊型病毒性肝炎(简称戊肝)和未分型病毒性肝炎(简称未分型肝炎)5种亚型^[1]。病毒性肝

炎是全球数百万慢性感染和死亡的罪魁祸首,对公共卫生仍构成重大威胁,其流行病学变化对社会和卫生服务提供有重要影响^[2]。因此,准确预测肝炎发病率对于识别其流行趋势、早期预防和更好地制定政府战略规划至关重要^[3]。目前,季节性自回归积分滑动平均模型(seasonal autoregressive integrated moving average, SARIMA)和神经网络模型被广泛应用于疾病发病率预测^[4-7]。作为基于径向基函数网络的神经网络模型,广义回归神经网络(generalized regression neural network, GRNN)模型相比 SARIMA 模型,在处

基金项目:重庆市公共卫生重点专科(学科)建设项目(YWBF2023081)

作者简介:孙亚军(1978—),男,硕士,主任医师,研究方向:疾病监测与数据分析

通信作者:姚远, E-mail: 71165057@qq.com

理复杂非线性关系、数据适应性方面具有显著优势,且预测精度高、训练快速、易于实现^[5,8-9]。2019 年,Francisco 等人^[10]基于 R 语言和 GRNN 神经网络,开发了用于时间序列预测的 tsfgrnn 包。目前,国内基于单一 GRNN 模型预测我国传染病发病率的报道较少^[11]。本文利用 2004—2019 年全国逐月病毒性肝炎发病率数据,探讨基于 R 语言 GRNN 模型在病毒性肝炎发病率预测中的应用,并与 SARIMA 模型(R 语言自动化建模)比较,以评估两种模型的拟合和预测效果,为 GRNN 模型在传染病日常监测预测预警中的应用和推广提供参考。

1 资料与方法

1.1 资料来源 2004—2019 年全国病毒性肝炎数据来源于公共卫生科学数据中心(<http://www.phsciencedata.cn/>),获取病毒性肝炎及各亚型逐月发病率数据。以 2004 年 1 月—2019 年 6 月数据拟合模型,并预测 2019 年 7—12 月发病率。

1.2 方法

1.2.1 GRNN GRNN 作为径向基(radial basis function, RBF)神经网络的一种变体,具有快速、单遍学习的特点^[10]。GRNN 具有 3 层,包括输入层、隐含层和输出层^[12]。隐含层是径向基层,基函数通常采用高斯函数: $G(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2})$,其中 x 为输入向量, x_i 为训练样本, σ 为平滑参数。GRNN 模型先通过隐含层产生一组权重 w_i (总和为 1,代表每个训练样本对最终结果的贡献),再通过输出层计算训练样本目标值 y_i 的加权平均值 \hat{y} ,作为 GRNN 输出向量。在 GRNN 模型中,平滑参数 σ 是一个关键参数,它控制着在计算 \hat{y} 时各个训练样本目标值的重要性。当 σ 较大时,所有训练样本目标值权重相似,导致模型结果趋于所有训练样本目标值的平均值;当 σ 较小时,最接近输入向量的训练样本目标值有显著的权重,使模型结果更倾向于这些样本的目标值^[10,13]。

R 语言中通过 tsfgrnn 包 grnn_forecasting 函数对时间序列数据建立 GRNN 模型,自回归滞后期(lags)、平滑参数(sigma)、多步向前预测策略(msas)、训练样本转换(transform)等是建模的关键设置^[10]。(1)自回归滞后期。该参数用于创建训练样本的输入向量,可手动设置,也可使用自动选择功能。自动选择基于以下标准进行。滞后期为 1 : f,其中 f 为时间序列的周期数。例如,季度数据的滞后期为 1 : 4,月度数据的滞后期为 1 : 12。对于周期数为 1 的时间序列(如年度数据),选择偏自相关函数(partial autocorrelation function, PACF)中具有显著自相关性

的滞后期。如果没有滞后期具有显著的自相关性,则选择滞后期为 1:5。(2)平滑参数 sigma 选取。如前所述,GRNN 对平滑参数非常敏感,因此为它选择合适的值至关重要。为了做出更好的选择,Francisco^[12]开发了滚动原点技术(rolling origin technique)探寻最优参数 sigma。该方法将历史数据分为训练数据和验证数据,逐步滚动进行预测,通过最小化平均绝对百分比误差(mean absolute percentage error, MAPE)来选择 sigma,从而使预测值与验证数据的误差最小。(3)多步向前预测策略。GRNN 预测策略包括多输入多输出策略(multiple-input multiple-output, MIMO)和递归策略(recursive),Francisco 等^[12]证实递归策略效果更好且速度更快,故参数默认设置为 recursive。(4)训练样本变换。如果时间序列数据具有趋势,则建议使用。该参数默认设置为加性变换(additive),适用于线性趋势的数据,对于指数性趋势的数据应设为乘法变换(multiplicative),对于没有明显趋势的数据,则设为不变换(none)。

1.2.2 SARIMA SARIMA 模型是 ARIMA 模型的扩展,是经典的时间序列模型之一,专门用于处理具有季节性波动的时间序列数据,SARIMA 模型结合了 ARIMA 模型和季节性成分,能够更好地捕捉和预测数据中的季节性模式,在疾病预测领域得到了广泛应用。SARIMA 原理见参考文献^[14]。在本研究中,利用 R 语言“forecast”包中 auto.arima()函数建立 SARIMA 模型^[15]。该函数能够自动选择最优的模型参数,包括自回归阶数(p)、差分次数(d)、滑动平均阶数(q)以及对应的季节性参数(P、D、Q)。其模型选择步骤如下:首先,选择差分次数 d 和 D。通过差分使序列平稳,再通过单位根检验(默认为 KPSS 检验,也可指定使用 ADF 或 PP 检验)选择合适的 d 和 D;其次,选择 AR 和 MA 阶数。通过评估不同的 p、q、P、Q 组合,基于最小化信息准则(默认为修正的赤池信息准则 AICc,也可指定使用 AIC 或 BIC)选择最优组合;最后,对模型残差进行白噪声检验(Ljung-Box 检验)以确保模型的适用性。

1.3 统计分析及模型评价 基于 R 语言“tidyverse”“forecast”“tsfgrnn”扩展包进行数据预处理、模型建立和模型预测。利用 ts()函数将原始数据转化为时间序列格式,采用 decompose()函数分解时间序列,并提取季节指数,该函数首先使用移动平均方法提取时间序列的趋势(trend)成分,再通过计算实际数据和趋势成分之间的差异,来识别和提取季节性成分(seasonal),最后通过将原始数据减去趋势和季节性成分,得到不规则成分(irregular);采用 auto.arima()函数自动识别并建立 SARIMA 模型,使用 forecast()函数对未来的

数据进行预测；通过 grnn_forecasting () 函数建立 GRNN 模型并进行预测。应用 checkresiduals() 函数进行 Ljung-Box 检验，以评估模型残差的自相关性，确保模型的适用性。模型拟合及预测效果评价采用平均绝对百分比误差 (MAPE)、平均绝对误差 (mean absolute error, MAE)、均方根误差 (root mean squared error, RMSE) 和平均误差率 (mean error rate, MER) 4 个指标。使用平均年度变化百分比 (average annual percent change, AAPC) 描述全国病毒性肝炎及各亚型发病率变化，采用 joinpoint 回归模型分析发病率时间趋势。所有检验均为双侧检验，检验水准为 0.05。

2 结果

2.1 疫情概况 2004—2019 年全国累积报告病毒性肝炎 20 777 873 例，年均发病率为 96.63/10 万。其中乙肝报告发病率最高，为 76.96/10 万，戊肝发病率最低，为 1.80/10 万。2004—2019 年全国病毒性肝炎和乙肝的发病率总体上无显著变化，AAPC 分别为

0.4% (95%CI: -0.9% ~ 1.7%) 和 0.3% (95%CI: -1.1% ~ 1.7%)。其中，2004—2007 年两者发病率均显著上升，AAPC 分别为 8.0% (95%CI: 2.6% ~ 13.6%) 和 9.4% (95%CI: 3.6% ~ 15.6%)；而在 2007—2015 年，两者发病率均显著下降，AAPC 分别为 -2.5% (95%CI: -3.8% ~ -1.2%) 和 -3.6% (95%CI: -5.0% ~ -2.2%)。丙肝和戊肝的发病率总体呈上升趋势，AAPC 分别为 11.6% (95%CI: 10.9% ~ 12.3%) 和 3.6% (95%CI: 1.9% ~ 5.3%)；其中，2004—2012 年两种肝炎的发病率快速增长，2012 年后增长速度放缓。甲肝和未分型肝炎的发病率总体呈下降趋势，AAPC 分别为 -11.4% (95%CI: -13.7% ~ -9.0%) 和 -12.0% (95%CI: -12.8% ~ -11.1%)；其中，2012—2019 年甲肝的下降趋势减缓；而未分型肝炎在 2014—2019 年呈加速下降趋势。分季节来看，病毒性肝炎、乙肝有明显的季节性；甲肝、丙肝、戊肝和未分型肝炎季节性不明显。见表 1，图 1 和 2。

表 1 2004—2019 年全国病毒性肝炎发病情况

Table 1 The annual incidence of viral hepatitis in China from 2004 to 2019

年份(年)	病毒性肝炎		甲肝		乙肝		丙肝		戊肝		未分型肝炎	
	发病数	发病率 (1/10 万)	发病数	发病率 (1/10 万)	发病数	发病率 (1/10 万)	发病数	发病率 (1/10 万)	发病数	发病率 (1/10 万)	发病数	发病率 (1/10 万)
2004	1 152 874	88.69	93 587	7.20	916 426	70.50	39 381	3.03	16 444	1.27	87 036	6.70
2005	1 195 355	91.96	73 349	5.64	982 297	75.57	52 927	4.07	15 541	1.20	71 241	5.48
2006	1 334 859	102.09	68 667	5.25	1 109 130	84.82	70 681	5.41	19 007	1.45	67 374	5.15
2007	1 425 428	108.44	77 135	5.87	1 169 946	89.01	92 378	7.03	20 577	1.57	65 392	4.97
2008	1 407 664	106.54	56 052	4.24	1 169 569	88.52	108 446	8.21	18 525	1.40	55 072	4.17
2009	1 425 020	107.30	43 841	3.30	1 179 607	88.82	131 849	9.93	20 275	1.53	49 448	3.72
2010	1 317 982	98.74	35 277	2.64	1 060 582	79.46	153 039	11.47	23 682	1.77	45 402	3.40
2011	1 372 344	102.34	31 456	2.35	1 093 335	81.54	173 872	12.97	29 202	2.18	44 479	3.32
2012	1 380 800	102.48	24 453	1.81	1 087 086	80.68	201 622	14.96	27 271	2.02	40 368	3.00
2013	1 251 872	92.46	22 244	1.64	962 974	71.12	203 155	15.00	27 902	2.06	35 597	2.63
2014	1 223 021	90.25	25 969	1.92	935 702	69.05	202 803	14.97	26 988	1.99	31 559	2.33
2015	1 218 946	89.47	22 667	1.66	934 215	68.57	207 897	15.26	27 169	1.99	26 998	1.98
2016	1 221 479	89.11	21 285	1.55	942 268	68.74	206 832	15.09	27 922	2.04	22 761	1.66
2017	1 283 523	93.02	18 875	1.37	1 001 952	72.61	214 023	15.51	29 014	2.10	19 284	1.40
2018	1 280 015	92.15	16 196	1.17	999 985	71.99	219 375	15.79	28 603	2.06	15 500	1.12
2019	1 286 691	92.13	19 271	1.38	1 002 292	71.77	223 660	16.02	28 155	2.02	12 961	0.93
合计	20 777 873	96.63	650 324	3.02	16 547 366	76.96	2 501 940	11.64	386 277	1.80	690 472	3.21

2.2 GRNN 模型建立 病毒性肝炎发病率数据为月度数据，根据 GRNN 模型输入向量数据选取规则，选取滞后后期 1:12 作为输入，sigma 参数设为 "ROLLING"，多步预测法选择递归法 ("recursive")，控制趋势选择加性变换 ("additive")，预测 2019 年 7—12 月发病率。通过滚动原点 ("rolling origin") 评估方法来评估模型拟合的准确性。拟合的最优模型参数及拟合的平均绝对百分比误差 MAPE，见表 2。

表 2 病毒性肝炎发病率 GRNN 最优模型及拟合效果

Table 2 Optimal GRNN models and their modeling accuracy for viral hepatitis incidence

肝炎类型	sigma	MAPE(%)
病毒性肝炎	0.78	1.80
甲肝	0.02	21.22
乙肝	0.28	2.10
丙肝	0.07	1.67
戊肝	0.03	4.43
未分型肝炎	0.01	11.43

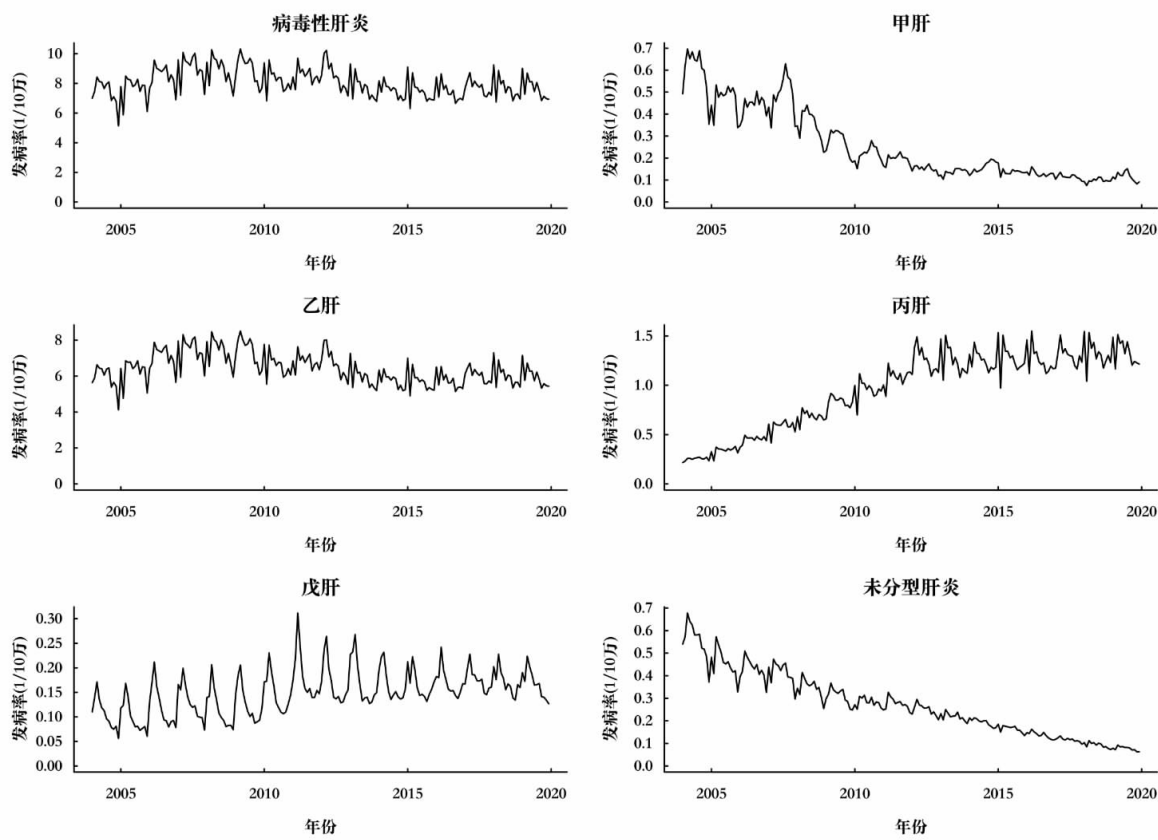


图 1 2004—2019 年全国病毒性肝炎发病率时间序列

Figure 1 The time series of viral hepatitis incidence rates in China from 2004 to 2019

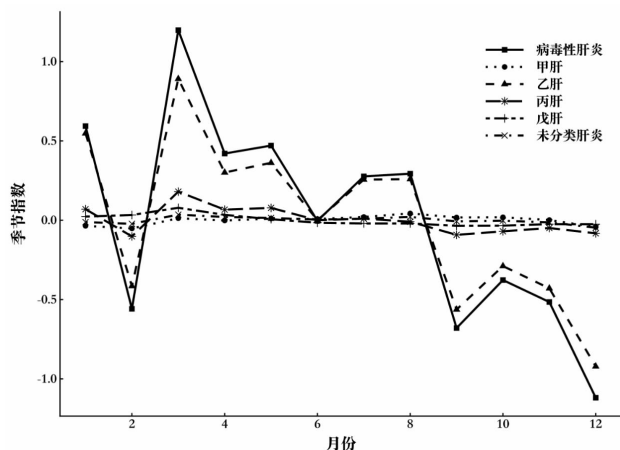


图 2 2004—2019 年全国病毒性肝炎季节指数

Figure 2 Seasonal index of viral hepatitis incidence rates in China from 2004 to 2019

2.3 SARIMA 模型建立 利用“forecast”包中的 auto.arima()函数自动建模,建模结果及残差自相关检验结果见表 3。病毒性肝炎、甲肝、乙肝、丙肝、戊肝的最优模型残差均为白噪声,序列信息提取完全,模型可以用于预测;未分类肝炎的最优模型残差可能存在轻微的自相关性,模型可以用于短期预测,但需关注预测误差。

2.4 两种模型拟合及预测效果比较 GRNN 模型的拟合 MAPE、MAE、RMSE、MER 分别介于 1.67% ~

表 3 病毒性肝炎发病率 SARIMA 最优模型及残差检验

Table 3 The optimal SARIMA models and residual test results for viral hepatitis incidence

肝炎类型	最优模型	Q^*	P 值
病毒性肝炎	SARIMA(2,1,0)(0,1,2) ₁₂	10.125	0.430
甲肝	SARIMA(1,1,3)(2,1,1) ₁₂	12.874	0.231
乙肝	SARIMA(2,1,0)(0,1,2) ₁₂	8.809	0.550
丙肝	SARIMA(2,1,1)(0,1,2) ₁₂	9.485	0.487
戊肝	SARIMA(1,0,2)(0,1,1) ₁₂	4.675	0.912
未分型肝炎	SARIMA(2,1,2)(2,1,1) ₁₂	17.005	0.074

21.22%、0.01 ~ 0.15、0.01 ~ 0.21、1.71% ~ 21.84%,而 SARIMA 模型分别介于 3.84 ~ 7.87、0.01 ~ 0.31、0.01 ~ 0.46、3.86% ~ 6.48%,GRNN 模型拟合误差波动幅度总体高于 SARIMA 模型,见表 4。GRNN 模型的预测 MAPE、MAE、RMSE、MER 分别介于 2.26 ~ 17.17、0.01 ~ 0.21、0.01 ~ 0.23、2.33% ~ 15.87%,而 SARIMA 模型分别介于 1.78% ~ 48.89%、0 ~ 0.19、0 ~ 0.21、1.83% ~ 43.77%,GRNN 模型的预测误差波动幅度总体小于 SARIMA 模型,见表 5。一般而言,模型的预测 MAPE 在 0 ~ 10%之间表示高预测精度,10% ~ 20%之间表示良好预测精度^[16]。GRNN 模型对于病毒性肝炎及各亚型预测的 MAPE 均在 20%以内,显示出良好的预测精度。

表 4 GRNN 模型与 SARIMA 模型拟合效果比较

Table 4 Comparison of modeling accuracy between GRNN and SARIMA models for viral hepatitis incidence

肝炎类型	GRNN				SARIMA			
	MAPE(%)	MAE	RMSE	MER(%)	MAPE(%)	MAE	RMSE	MER(%)
病毒性肝炎	1.80	0.15	0.21	1.84	3.84	0.31	0.46	3.86
甲肝	21.22	0.03	0.03	21.84	7.87	0.02	0.03	6.48
乙肝	2.10	0.13	0.18	2.14	3.90	0.25	0.37	3.93
丙肝	1.67	0.02	0.03	1.71	4.74	0.05	0.07	4.76
戊肝	4.43	0.01	0.01	4.39	5.94	0.01	0.01	6.19
未分型肝炎	11.43	0.01	0.01	11.51	4.71	0.01	0.02	4.19

表 5 GRNN 模型与 SARIMA 模型预测效果比较

Table 5 Comparison of the accuracy of viral hepatitis incidence forecasts using GRNN and SARIMA models

肝炎类型	GRNN				SARIMA			
	MAPE(%)	MAE	RMSE	MER(%)	MAPE(%)	MAE	RMSE	MER(%)
病毒性肝炎	2.91	0.21	0.23	2.94	2.54	0.19	0.21	2.58
甲肝	17.17	0.02	0.02	15.87	48.89	0.05	0.05	43.77
乙肝	2.26	0.13	0.16	2.33	1.78	0.10	0.13	1.83
丙肝	2.28	0.03	0.04	2.37	3.42	0.04	0.06	3.47
戊肝	8.85	0.01	0.02	8.24	11.88	0.02	0.02	10.77
未分型肝炎	8.41	0.01	0.01	8.26	3.27	0.00	0.00	3.19

3 讨论

本文基于 R 语言 tsfgrnn 包介绍 GRNN 模型在病毒性肝炎及其亚型发病率预测中的应用。研究结果表明,GRNN 模型具有较高的预测精度;与 SARIMA 模型相比,GRNN 模型预测病毒性肝炎及其亚型发病率更为稳健,这与既往同类研究结果相似^[9,17-18]。Francisco 等曾比较 GRNN 模型与指数平滑模型、自动化 SARIMA 模型、K 近邻时间序列预测模型、神经网络自回归模型和多层感知器模型的预测效果,结果显示 GRNN 模型优于 K 近邻时间序列预测模型、神经网络自回归模型、多层感知器模型,但略逊于指数平滑模型、自动化 SARIMA 模型^[10]。然而,中国学者研究显示,在肾综合征出血热和新型冠状病毒肺炎发病率预测中,GRNN 模型预测误差远高于 SARIMA 模型^[19-20]。

本研究中,就预测误差 MAPE 而言,GRNN 模型预测病毒性肝炎、乙肝、未分型肝炎发病率误差均高于 SARIMA 模型,但对于其他三种亚型的肝炎,情况刚好相反。这种以 SARIMA 模型相比较的 GRNN 模型发病率预测误差高低不一的原因,可能与 GRNN 模型过拟合、不同的疾病特征、地区与监测方法差异、数据自身特性及模型适用性有关^[18,20]。研究表明,GRNN 模型在捕捉非线性关系时表现优越,而 SARIMA 模型在处理线性和季节性数据时更为有效^[17,21]。有学者利用两者优势,发展出 SARIMA-GRNN 组合模型,其发病率预测效果优于单一的 SARIMA 模型^[22-23]。基于 R 语言的 SARIMA-GRNN 组合自动化模型对病毒性肝炎及其亚型发病率的预测效果如何,尚有待于后续进一步研究。

作为神经网络家族的一员,GRNN 模型本质上也是一种“黑箱”技术,因此,该模型依然存在专业解释性差的问题^[24]。然而,与既往研究中神经网络预测涉及大量参数,缺乏系统化建模流程和策略不同^[25],基于 R 语言 tsfgrnn 包的 GRNN 模型已经以自动化方式整合了最佳建模方法和训练样本转换策略^[12]。因此,相比 SARIMA 模型,本文介绍的 GRNN 模型具有以下优点^[5,8,12]:(1)非线性关系的捕捉:GRNN 模型能更好地捕捉时间序列中的非线性关系。(2)建模快速与易用性:GRNN 模型通过单次训练并仅需调整一个平滑参数来实现快速建模,并且不需要传统的统计建模经验。此外,tsfgrnn 包提供了自动选择最优参数的功能,使其在快速自动化预测中更加便捷。尽管 forecast 包提供了自动选择 SARIMA 模型参数的功能,但这些参数仍需通过统计检验才能最终确定,因此 SARIMA 模型的自动化流程通常仍需专家指导。(3)稳健性与适应性:GRNN 模型通过平滑参数调整和训练样本转换来适应不同的数据模式,从而在变化的数据环境中保持较高的预测精度和灵活性。

综上所述,GRNN 为一种无需统计建模经验、操作简单、功能强大的时间序列预测模型,对传染病发病率时间序列的预测效果较好,值得进一步推广应用。本研究也存在一定的局限性^[17,20,26-27]:一是,未能对更密集的日、周数据及不同流行季节数据进行预测效果分析;二是,模型仅考虑时间因素,属于单因素模型,无法加入其它疾病影响因素,如气候、重大事件等,与复杂的病毒性肝炎流行可能不相适应。

利益冲突声明 本研究不存在任何利益冲突

参考文献

- [1] Usuda D, Kaneoka Y, Ono R, et al. Current perspectives of viral hepatitis [J]. *World Journal of Gastroenterology*, 2024, 30(18): 2402–2417.
- [2] Black AP, Wallace J, Binka M, et al. Correction: the challenges of viral hepatitis elimination: a global response to a global problem[J]. *BMC Public Health*, 2023, 23(1): 1229.
- [3] Xia ZH, Qin L, Ning Z, et al. Deep learning time series prediction models in surveillance data of hepatitis incidence in China [J]. *PLOS One*, 2022, 17(4): e0265660.
- [4] Zhai MM, Li WH, Tie P, et al. Research on the predictive effect of a combined model of ARIMA and neural networks on human brucellosis in Shanxi Province, China: a time series predictive analysis[J]. *BMC Infectious Diseases*, 2021, 21(1): 280.
- [5] Li ZQ, Wang ZZ, Song H, et al. Application of a hybrid model in predicting the incidence of tuberculosis in a Chinese population[J]. *Infection and Drug Resistance*, 2019, 12: 1011–1020.
- [6] Feng Y, Cui X, Lv J, et al. Deep learning models for hepatitis E incidence prediction leveraging meteorological factors[J]. *PLOS One*, 2023, 18(3): e0282928.
- [7] Mollalo A, Rivera KM, Vahedi B. Artificial neural network modeling of novel coronavirus (COVID–19) incidence rates across the continental United States [J]. *International Journal of Environmental Research and Public Health*, 2020, 17(12): 4204.
- [8] Zhu XX, Zheng YM, Zheng YL, et al. Application of generalized regression neural network and sarima model in prediction of AIDS [EB/OL]. [2024–10–20]. <https://ieeexplore.ieee.org/document/9988161>.
- [9] 冯佳宁, 肖宇飞, 王晓璇, 等. ARIMA 乘积季节模型与 GRNN 模型在猩红热发病预测中的比较[J]. *中国卫生统计*, 2020, 37(6): 851–854.
- Feng JN, Xiao YF, Wang XX, et al. A comparative study of multiplicative seasonal ARIMA model and GRNN for scarlet fever incidence forecasting [J]. *Chinese Journal of Health Statistics*, 2020, 37(6): 851–854.(In Chinese)
- [10] Martínez F, Charte F, Rivera AJ, et al. Automatic time series forecasting with GRNN: A comparison with other models[EB/OL]. [2024–10–20]. https://link.springer.com/chapter/10.1007/978-3-030-20521-8_17.
- [11] 李于于, 曾令城, 柳巍. 遗传优化的广义回归神经网络在肺结核发病预测中的应用 [J]. *现代预防医学*, 2020, 47(21): 3863–3866, 3923.
- Li YY, Zeng LC, Liu W. Application of genetic-optimized generalized regression neural network in the prediction of tuberculosis [J]. *Modern Preventive Medicine*, 2020, 47(21): 3863–3866, 3923.(In Chinese)
- [12] Martínez F, Charte F, Frías MP, et al. Strategies for time series forecasting with generalized regression neural networks [J]. *Neurocomputing*, 2022, 491: 509–521.
- [13] 汪宗正. 基于广义回归神经网络的动态光散射粒度反演研究 [D]. 淄博: 山东理工大学, 2023.
- Wang ZZ. Study on particle size distribution inversion of dynamic light scattering based on generalized regression neural network [D]. Zibo: Shandong University of Technology, 2023.(In Chinese)
- [14] Box G, Jenkins GM, Reinsel GC, et al. *Time series analysis: forecasting and control*(5th ed.)[M]. New Jersey: John Wiley&Sons, 2015.
- [15] Hyndman R, Athanasopoulos G, Bergmeir C, et al. 2024. Forecast: forecasting functions for time series and linear models [EB/OL]. [2024–10–20]. <https://pkg.robjhyndman.com/forecast/>.
- [16] Dun M, Xu ZC, Chen Y, et al. Short-term air quality prediction based on fractional grey linear regression and support vector machine[J]. *Mathematical Problems in Engineering*, 2020, 249: 1–13.
- [17] Wei WD, Jiang JJ, Liang H, et al. Application of a combined model with autoregressive integrated moving average (ARIMA) and generalized regression neural network (GRNN) in forecasting hepatitis incidence in Heng county, China [J]. *PLOS One*, 2016, 11(6): e0156768.
- [18] Wang YW, Shen ZZ, Jiang Y. Comparison of autoregressive integrated moving average model and generalised regression neural network model for prediction of haemorrhagic fever with renal syndrome in China: a time-series study[J]. *BMJ Open*, 2019, 9(6): e025773.
- [19] 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(4): 548–554.
- Liu T, Yao ML, Hou QB, et al. Comparison of seven time series models in fitting and predicting the incidence of hemorrhagic fever with renal syndrome in China [J]. *Chinese Journal of Vector Biology and Control*, 2022, 33(4): 548–554.(In Chinese)
- [20] Wang G, Wu T, Wei W, et al. Comparison of ARIMA, ES, GRNN and ARIMA-GRNN hybrid models to forecast the second wave of COVID–19 in India and the United States [J]. *Epidemiology and Infection*, 2021, 149: e240.
- [21] Guo YH, Feng Y, Qu FL, et al. Prediction of hepatitis E using machine learning models[J]. *PLOS One*, 2020, 15(9): e0237750.
- [22] 李峰, 陈国清, 徐士林, 等. 三种时间序列模型在盐城市肾综合征出血热发病预测应用中的比较研究 [J]. *中国卫生统计*, 2020, 37(2): 249–252.
- Li F, Chen GQ, Xu SL, et al. Comparative analysis of three time series models for predicting hemorrhagic fever with renal syndrome incidence in Yancheng City [J]. *Chinese Journal of Health Statistics*, 2020, 37(2): 249–252.(In Chinese)
- [23] 李文豪, 曾显兴, 李笑颜, 等. SARIMA-GRNN 组合模型在伤寒与副伤寒逐月发病数预测中的应用 [J]. *中华疾病控制杂志*, 2021, 25(11): 1341–1346.
- Li WH, Zeng YX, Li XY, et al. Application of SARIMA-GRNN combined model in forecasting the monthly incidence of typhoid fever and paratyphoid fever [J]. *Chinese Journal of Disease Control & Prevention*, 2021, 25(11): 1341–1346.(In Chinese)
- [24] Burkart N, Huber MF. A survey on the explainability of supervised machine learning [J]. *Journal of Artificial Intelligence Research*, 2021, 70: 245–317.
- [25] Du XP, Xu HY, Zhu F. Understanding the effect of hyperparameter optimization on machine learning models for structure design problems[J]. *Computer Aided Design*, 2021, 135: 103013.
- [26] Yan D, Kong Y, Ye B, et al. Spatio-temporal variation and daily prediction of PM_{2.5} concentration in world-class urban agglomerations of China[J]. *Environmental Geochemistry and Health*, 2021, 43(1): 301–316.
- [27] Wang L, Xia Y, Lu YC. A novel forecasting approach by the GA-SVR-GRNN hybrid deep learning algorithm for oil future prices [J]. *Computational Intelligence and Neuroscience*, 2022(1): 4952215.