

感 染

DOI:10.13406/j.cnki.cyxh.001542

重庆市手足口病多种预警模型的建立与评估

何晶晶,明鑫,刘勋,谢佳伽,王怡,邓丹

(重庆医科大学公共卫生与管理学院卫生统计与信息管理教研室,重庆 400016)

【摘要】目的:构建多种重庆市手足口病(hand, foot and mouth disease, HFMD)疫情预测模型并评价模型的拟合和预测效果,为卫生部门提供准确的疾病动态及预警预告奠定基础。**方法:**对重庆市 2009 至 2014 年 HFMD 月发病率资料分别建立 3 种单一预测模型(SARIMA、BPNN 和 Elman 神经网络)和 3 种组合预测模型(SARIMA-BPNN、SARIMA-Elman、BPNN-Elman),利用 2015 年 HFMD 月发病率资料对各个模型进行拟合和评价,并选择最优预警模型预测 2016 至 2017 年重庆市 HFMD 发病率。**结果:**SARIMA 模型、BPNN 模型、Elman 模型、SARIMA-BPNN 组合模型、SARIMA-Elman 组合模型、BPNN-Elman 组合模型拟合及预测的平均绝对误差(mean absolute error, MAE)、平均误差率(mean error rate, MPAE)、预测准确度(prediction accuracy, P)、非线性相关系数(non-linear correlation coefficient, RNL)和平均相对误差(mean relative error, MRE)分别为 0.595、0.515、0.485、0.551、0.493;0.265、0.229、0.771、0.673、0.029;0.234、0.202、0.798、0.737、0.142;0.248、0.215、0.785、0.714、0.101;0.328、0.284、0.716、0.620、0.103;0.229、0.198、0.802、0.745、0.071。综合各个评价指标,选出本次研究最优预测模型为 BPNN-Elman 组合模型,并采用 BPNN-Elman 组合模型预测出 2016 年和 2017 年重庆市 HFMD 年平均发病率分别为 162.8/10 万和 160.9/10 万,95%置信区间为-0.236~0.184。**结论:**BPNN-Elman 组合模型预测效果最优,能较好预测重庆市 HFMD 发病情况,为 HFMD 防控提供科学依据。

【关键词】手足口病;预测;组合模型;ARIMA 模型;BP 神经网络;Elman 神经网络

【中图分类号】R181.2

【文献标志码】A

【收稿日期】2018-01-22

Establishment and evaluation of multiple warning models for hand, foot and mouth disease in Chongqing

He Jingjing, Ming Xin, Liu Xun, Xie Jiajia, Wang Yi, Deng Dan

(Teaching and Research Section of Health Statistics and Information Management,
School of Public Health and Management, Chongqing Medical University)

【Abstract】Objective: To lay the foundation for the health sector and to provide accurate disease dynamics and early warning by constructing a variety of prediction models of hand, foot and mouth disease (HFMD) in Chongqing and to evaluate its effect. **Methods:** Three kinds of single prediction models (SARIMA, BPNN and Elman neural network) and three combined prediction models (SARIMA-BPNN, SARIMA-Elman, BPNN-Elman) were established for the monthly incidence of HFMD in Chongqing from 2009 to 2014. The models were evaluated by using the monthly incidence data of HFMD in 2015, and the optimal warning model was selected to predict the incidence of HFMD in Chongqing from 2016 to 2017. **Results:** The results showed that the mean absolute error (MAE), mean error rate (MPAE), prediction accuracy (P), non-linear correlation coefficient (RNL) and mean relative error (MRE) of the SARIMA model, BPNN model, Elman model, SARIMA-BPNN, SARIMA-Elman, BPNN-Elman combined model respectively were 0.595, 0.515, 0.485, 0.551, 0.493; 0.265, 0.229, 0.771, 0.673, 0.029; 0.234, 0.202, 0.798, 0.737, 0.142; 0.248, 0.215, 0.785, 0.714, 0.101; 0.328, 0.284, 0.716, 0.620, 0.103; 0.229, 0.198, 0.802, 0.745, 0.071. Through comprehensive evaluation of various indicators, the optimal prediction model is BPNN-Elman combination model. The average annual incidence of HFMD of the year 2016 and the year 2017 in Chongqing were 162.8/10 million and 160.9/10 million based on BPNN-Elman combination model with 95% confidence interval of -0.236 to 0.184. **Conclusion:** The BPNN-Elman combined model has the best prediction effect and can better predict the

作者介绍:何晶晶, Email: jing_brush@126.com,

研究方向:人群健康与统计决策。

通信作者:邓丹, Email: ddlinger@126.com。

基金项目:重庆市医学科研计划资助项目(编号:2015MSXM094)。

优先出版: <http://kns.cnki.net/kcms/detail/50.1046.R.20180321.1618.048.html>
(2018-03-22)

incidence of HFMD in Chongqing, which provides a scientific basis for prevention and treatment of HFMD.

【Key words】 hand, foot and mouth disease; prediction; combination model; ARIMA model; BP neural network; Elman neural network

手足口病(hand, foot and mouth disease, HFMD)是由多种人肠道病毒引起的一种儿童常见传染病,也是我国法定报告管理的丙类传染病。我国自 1981 年在上海始见 HFMD,之后全国多地均有报道,且以广东、重庆和上海等地尤为多发^[1-2]。近几年重庆市 HFMD 发病率居高不下,其中 2014 年发病率 163.95/10 万,2015 年发病率 138.711/10 万,高于同期河北省^[3]、贵州省^[4]、唐山市^[5]、呼伦贝尔市^[6]等地,略低于同期全国 HFMD 平均发病率。鉴于 HFMD 在重庆市的流行现况依然不容乐观,建立高效、准确、合理的 HFMD 预警模型,并根据预警信号及早采取预防和控制措施,以期减少 HFMD 的发病率和死亡率是必要的。

目前, HFMD 预警模型种类众多,主要有差分自回归移动平均模型(autoregressive integrated moving average model, ARIMA)、空间统计学、线性回归模型、Elman 神经网络模型、BP 神经网络模型(back propagation neural network, BPNN)等。其中,以 ARIMA 模型、BPNN 模型较为常用。有研究显示 Elman 神经网络的预测效果好于 BPNN 和 ARIMA 模型^[7]。不过,大部分研究者运用单一预测模型研究较多,而组合预测模型研究较少。所谓组合预测模型就是将不同的单一预测模型按一定方式进行组合,综合利用各种方法所能提供的信息,从而达到提高预测精度和增加预测稳定性的效果^[8]。吴文博等^[9]利用遗传算法优化的 ARIMA-BPNN 组合模型预测 HFMD 发病趋势,结果表明组合预测模型比单一预测模型预测精度更高,效果更好。

因此,本研究拟分别采用单一预测模型(SARIMA、BPNN 和 Elman 神经网络)和组合预测模型(SARIMA-BPNN、SARIMA-Elman、BPNN-Elman)对重庆市月发病数据进行建模与预测,评价几种模型的预测效果,并选择最优预警模型预测 2016 至 2017 年重庆市 HFMD 发病率,以期发现适合重庆市 HFMD 发病率的预测模型,为 HFMD 及其他传染病发病率的监测提供新思路。

1 材料与方法

1.1 资料来源

重庆市 2009 至 2015 年 HFMD 发病资料来源于重庆市疾病预防控制中心传染病直报系统, HFMD 病例分别由各区县的疾病预防控制中心确诊。重庆市人口学资料来源于《重庆市卫生统计年鉴》。因 HFMD 具有明显的季节性和周期

性,故将 HFMD 发病数据按月份整理为月发病率数据。

1.2 研究方法

基于重庆市 2009 至 2014 年 HFMD 月发病数据建立 3 种单一预测模型(SARIMA、BPNN、Elman)和 3 种组合预测模型(SARIMA-BPNN、SARIMA-Elman、BPNN-Elman),用 2015 年重庆市 HFMD 月发病数据进行各个模型的拟合和评价。因此将 2009 年 1 月至 2014 年 12 月重庆市 HFMD 月发病率数据作为各个预测模型的训练集,2015 年 1 月至 2015 年 12 月的月发病率数据作为测试集建立以上 6 种预警模型。并采用平均绝对误差(mean absolute error, MAE)、平均误差率(mean error rate, MAPE)、预测准确度(prediction accuracy, P)、非线性相关系数(non-linear correlation coefficient, RNL)和平均相对误差(mean relative error, MRE)等指标评价以上 6 种模型的预测效果。当 MAE、MAPE 和 MRE 越接近 0, P 和 RNL 越接近于 1, 模型预测效果越好。具体公式如下:

$$MAE = \frac{1}{12} \sum_{i=1}^{12} \text{abs}(p(i) - t(i))$$

$$MAPE = \frac{MAE}{\frac{1}{12} \sum_{i=1}^{12} t(i)} \times 100\%$$

$$P = (1 - MAPE)$$

$$RNL = 1 - \sqrt{\frac{\sum (p(i) - t(i))^2}{\sum p(i)^2}}$$

$$MRE = \frac{1}{12} \sum_{i=1}^{12} \frac{\text{abs}(t(i) - p(i))}{t(i)} \times 100\%$$

其中, $p(i)$ 代表预测值, $t(i)$ 代表实际值。

最后,选取最优预测模型对重庆市 2016 至 2017 年 HFMD 流行趋势作出预测。

1.3 模型基本原理

1.3.1 ARIMA 乘积季节性模型 ARIMA 模型是将非平稳时间序列转化为平稳时间序列,然后将因变量对它的滞后值以及随机误差项的现值和滞后值进行回归所建立的模型^[10],主要包含乘积季节性 ARIMA 模型(seasonal ARIMA, SARIMA)和非季节性 ARIMA 模型。由于时间序列常呈周期性变化(季节性趋势),用普通的 ARIMA 模型处理这种季节性趋势,会导致参数过多,模型复杂,而 SARIMA 模型可以得到参数简约的模型。

1.3.2 BP 神经网络模型 BP 神经网络是一种按误差逆传播算法训练的多层前馈网络,是目前应用最广泛的神经网络模型之一。它是人工神经网络的一种,通过模仿生物大脑神经系统信息处理功能实现输入与输出之间的任意非线性优化映射,有着传统统计方法无法比拟的适应性、容错性及自组织性等优点。

1.3.3 Elman 神经网络模型 Elman 神经网络是一种典型的非线性动态神经网络,具有自组织、自学习的特征。Elman 网

络一般分为 4 层,即输入层、隐含层、承接层和输出层。与 3 层的 BP 神经网络相比,Elman 增加了网络本身处理动态信息的能力,达到动态建模的目的^[11],近年来被用于传染病发病率预测。

1.3.4 组合预测模型 目前,HFMD 疫情单一预测模型研究较多,组合预测模型研究较少。但单一的预测模型有一定的局限,有的预测模型其预测结果比较滞后,有的预测结果却提前。如果将多种预测模型组合起来综合应用,那么就能摒弃各自的不足,汲取各自的优点,从而预测出更为准确的结果^[12]。较为常见的组合预测模型有 ARIMA-BPNN 的混合预测模型、ARIMA 结合泊松分布方法、ARIMA-RBF 组合模型等。由于 HFMD 发病率序列的拟合值与实际值存在高度的关联性,故将单一预警模型(A)的拟合值作为另一种预警模型(B)的输入,HFMD 实际月发病率作为 B 模型的输出,构建组合预警模型。本文选定的 A 和 B 模型为 SARIMA 模型、BPNN 模型和 Elman 神经网络模型,因此最终构建 3 种组合预警模型为 SARIMA-BPNN、SARIMA-Elman、BPNN-Elman。

1.4 统计分析

采用 SPSS 21.0 软件对重庆市 HFMD 疫情资料进行统计描述;通过 SAS 9.2 软件对重庆市各年发病率进行卡方趋势检验,并构建 SARIMA 预测模型;采用 MATLAB R2015b 软件构建 BP 神经网络模型、Elman 神经网络模型和 3 种组合预测模型(SARIMA-BPNN、SARIMA-Elman、BPNN-Elman)。

2 结果

2.1 重庆市 HFMD 发病概况

2009 至 2015 年重庆市共报告 211 416 例 HFMD 病例,其中男性 125 437 例,女性 85 979 例。2009 至 2014 年发病呈上升趋势,2015 年发病率较 2014 年略有下降,但仍高于其余年份,总体处于较高发病水平。发病率最高峰值出现在 2014 年(28.21/10 万)。经卡方趋势检验发现近 6 年发病率有统计学意义($Z=203.650, P=0.000$),如图 1 所示。

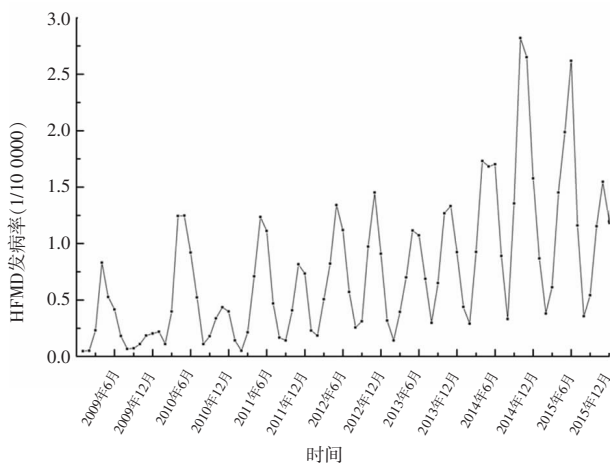


图 1 2009 年至 2015 年重庆市 HFMD 年报告发病率(1/10 万)

2.2 SARIMA

图 1 显示重庆市 HFMD 月发病率既含有长期趋势又含有以年为周期的季节效应,为不平稳序列。故将原始序列进行自然对数转换和一次季节差分,消除季节和趋势的影响,以获得稳定的方差和均值。差分后的序列为非白噪声序列,可以进行时间序列分析。根据 ACF 和 PACF 图,以及 X-12 模型的自动识别功能,选定 ARIMA 模型为(1,0,0)(0,1,0)₁₂。其标准化 BIC 值=0.842,在拟合的所有模型中最小,为最优模型。采用 Ljung-Box 方法检验残差白噪声,得到 Ljung-Box 统计量为 19.224, $P=0.316$ 。对模型残差进一步做 ACF 和 PACF 图可知该残差是白噪声序列,且模型中的参数检验有统计学意义。因此,ARIMA(1,0,0)(0,1,0)₁₂ 模型是最适合本次研究的预测模型。

利用建立的 ARIMA(1,0,0)(0,1,0)₁₂ 模型对重庆市 2015 年 HFMD 发病率进行拟合,绘制实际值和预测值的拟合图。由图 2 可见,该模型拟合结果的动态变化趋势与实际情况基本一致。计算出预测结果的 $MAE=0.595, MAPE=51.5\%, P=48.5\%, RNL=0.551, MRE=49.33\%$ 。

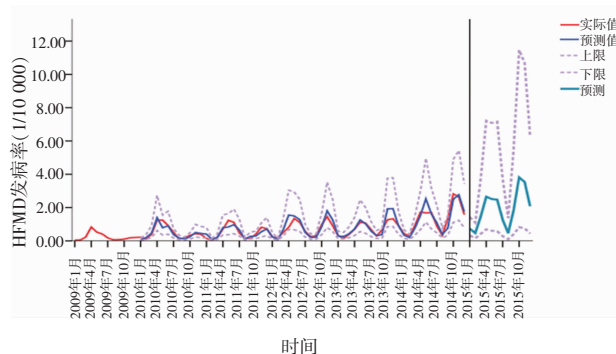


图 2 重庆市 2009 至 2015 年 HFMD 发病率实际值和预测值的拟合图

2.3 BP 神经网络

采用 2009 至 2015 年 HFMD 月发病率数据构建 4 个 BP 神经网络 net1~net4。分别为 net1:用前 6 个月 HFMD 月发病率预测后 1 个月发病率;net2:用前 12 个月发病率预测后 1 个月发病率;net3:用前 3 年历史同期发病率预测当前月份发病率;net4:用前 4 年历史同期发病率预测当前月份发病率。以构建 net4 为例,输入节点为 4,输出节点为 1。共构建 36 个样本,前 24 个样本为训练样本,后 12 个样本为测试样本。采用 mapminmax 函数对 HFMD 发病率进行归一化处理,将所有数据线性映射到 [-1,1] 区间。传递函数选择 tansig 函数和 purelin 函数,训练函数用 trainlm 函数。最大训练次数为 3 000,训练精度为 $1e-7$,学习速率为 0.01,显示训练迭代过程为 1 000。

根据隐含层节点参考公式选定隐含层范围为 3~13,分别训练样本数据集,选取训练样本拟合结果 MSE 最小的网

络作为最优网络。通过对神经网络模型网络的反复验证,当各个网络隐含层结点分别取 5、5、8、7 时,得到的 MSE 最小。用以上建立的 4 个最优网络分别对 2015 年 1 月到 12 月进行预测,计算预测的月发病率与实际月发病率之间的 MAE、MPAE、P、RNL 和 MRE(见表 1)。从表中可以看出这 4 个网络预测效果均较理想,但 net4 的 MAE、MAPE 和 MRE 值最小, P 值最大,且网络结构也较简单,因此选择 net4(4-7-1)为最优的 BP 神经网络模型。

表 1 4 个 BP 神经网络 HFMD 月发病数据拟合情况(1/10 000)

月份	实际值	拟合值			
		net1	net2	net3	net4
2015 年 1 月	0.869	1.036	0.261	0.801	0.503
2015 年 2 月	0.380	0.484	0.319	0.344	0.446
2015 年 3 月	0.613	0.588	1.212	1.016	0.861
2015 年 4 月	1.451	0.984	1.618	1.036	1.181
2015 年 5 月	1.987	1.887	1.333	2.335	1.727
2015 年 6 月	2.618	1.905	2.036	2.250	1.552
2015 年 7 月	1.160	1.183	0.653	1.103	1.005
2015 年 8 月	0.356	0.909	0.660	0.773	0.449
2015 年 9 月	0.542	0.403	1.091	0.458	0.884
2015 年 10 月	1.153	0.448	1.323	1.466	1.240
2015 年 11 月	1.547	1.442	1.235	2.309	1.479
2015 年 12 月	1.196	2.006	1.653	1.612	1.350
MAE		0.326	0.414	0.307	0.265
MPAE		0.282	0.358	0.266	0.229
P		0.718	0.642	0.734	0.771
RNL		0.650	0.629	0.747	0.673
MRE		0.091	0.120	0.193	0.029

2.4 Elman 神经网络模型

同 BP 神经网络的构建原理,采用 2009 至 2015 年 HFMD 月发病率数据构建 4 个 Elman 神经网络 net5~net8。隐含层节点数的确定一般没有特定的要求,本研究利用原始的疫情资料,分别建立 30(可调)个不同隐含层节点数的 Elman 神经网络模型,最大训练次数为 5 000,训练精度为 $1e-7$,学习速率为 0.01,显示训练迭代过程为 1 000,函数选择同 BPNN。当隐含层节点数分别为 13、9、23、18 时,4 个网络的 MSE 最小。分别用建立的 4 个 Elman 网络对 2015 年重庆市 HFMD 发病率进行拟合,拟合结果见表 2。从表 2 中可以看出 net8 的 MAE 和 MAPE 值最小, P 值最大,RNL 值也最接近 1,因此选择 net8(4-18-1-1)为最优的 Elman 神经网络模型。

2.5 组合预测模型及模型评价

利用已建立的 SARIMA 模型对重庆市 2009 至 2014 年 HFMD 发病率进行拟合,由于 SARIMA 模型对 HFMD 发病率序列的拟合值与实际值存在高度的关联性,考虑将 SARIMA 模型对 HFMD 的拟合值作为 BPNN 和 Elman 模型的输入,将对应月份 HFMD 实际发病率作为 BPNN 和 ELMAN 神经

网络的输出值,构建 1 维输入,1 维输出的 SARIMA-BPNN 和 SARIMA-Elman 组合预测模型,这 2 个组合模型中间层的节点数通过多次实验对比,分别最终选定为 5 和 6,其余参数设置同 BPNN 和 Elman 模型。同时,利用已建立的最优 BP 神经网络模型对 HFMD 进行拟合,拟合值作为 Elman 神经网络的输入值,实际发病率作为 Elman 网络的输出值,建立 BPNN-Elman 组合预测模型,当隐含层节点为 9 时,BPNN-Elman 模型的 MSE 值最小。

表 2 4 个 Elman 神经网络 HFMD 月发病数据拟合情况(1/10 000)

月份	实际值	拟合值			
		net5	net6	net7	net8
2015 年 1 月	0.869	0.522	0.641	0.585	0.927
2015 年 2 月	0.380	0.867	0.488	0.407	0.399
2015 年 3 月	0.613	1.043	0.981	0.859	0.880
2015 年 4 月	1.451	1.110	1.160	1.499	1.274
2015 年 5 月	1.987	1.497	1.766	1.615	1.747
2015 年 6 月	2.618	1.643	1.427	1.677	1.892
2015 年 7 月	1.160	1.490	1.260	1.189	1.517
2015 年 8 月	0.356	0.370	0.420	0.493	0.507
2015 年 9 月	0.542	0.719	0.961	1.504	0.795
2015 年 10 月	1.153	1.017	1.157	1.976	1.601
2015 年 11 月	1.547	1.682	1.591	1.852	1.542
2015 年 12 月	1.196	1.250	1.373	1.619	1.297
MAE		0.326	0.268	0.383	0.234
MPAE		0.282	0.232	0.331	0.202
P		0.718	0.768	0.669	0.798
RNL		0.650	0.655	0.633	0.763
MRE		0.116	0.089	0.257	0.142

用上述建立的 3 种组合模型对 2015 年重庆市 HFMD 发病率进行拟合,拟合结果详见表 3。由表 3 可知,BPNN-Elman 神经网络 MAE 和 MPAA 值最小, P 值和 RNL 值最大,且平均相对误差为 7.08%,用于预测时准确度和精度较高,提示 BPNN-Elman 神经网络预测准确性高于其余 5 个预测模型。综合以上 5 个评价指标,以预测准确度 P 值为代表,6 个预警模型预测效果从高到低依次为:BPNN-Elman>Elman>SARIMA-BPNN>BPNN>SARIMA-Elman>SARIMA。图 3 为 3 种组合预测模型对 HFMD 发病率的拟合曲线,BPNN-Elman 组合模型的拟合曲线与原始值最接近。

2.6 重庆市 2016 至 2017 年 HFMD 发病率预测

用 BPNN-Elman 组合模型对历史数据进行仿真,并对重庆市 2016 至 2017 年 HFMD 发病率作出预测,绘制预测拟合图(图 4)。从图 4 中可以看出,模型拟合的动态趋势与实际情况基本一致。预测 2016 年 HFMD 总发病率为 162.8/10 万,2017 年的总发病率为 160.9/10 万($95\%CI=-0.236-0.184$),均高于 2015 年发病率。其中 4 至 7 月和 10 至 12 月为发病高峰期。

表 3 6 个预测模型 HFMD 月发病数据拟合情况(1/10 000)

月份	实际值	拟合值					
		SARIMA	BPNN	Elman	SARIMA-BP	SARIMA-Elman	BPNN-Elman
2015 年 1 月	0.869	0.683	0.503	0.927	0.875	0.629	0.448
2015 年 2 月	0.380	0.420	0.446	0.399	0.358	0.403	0.351
2015 年 3 月	0.613	1.260	0.861	0.880	1.052	1.016	1.198
2015 年 4 月	1.451	2.234	1.181	1.274	1.654	1.538	1.271
2015 年 5 月	1.987	2.082	1.727	1.747	1.654	1.486	1.950
2015 年 6 月	2.618	2.032	1.552	1.892	1.654	1.466	2.033
2015 年 7 月	1.160	1.031	1.005	1.517	1.213	0.867	1.206
2015 年 8 月	0.356	0.375	0.449	0.507	0.327	0.358	0.351
2015 年 9 月	0.542	1.500	0.884	0.795	0.740	1.172	1.022
2015 年 10 月	1.153	3.071	1.240	1.601	1.654	1.607	1.205
2015 年 11 月	1.547	2.845	1.479	1.542	1.654	1.613	1.274
2015 年 12 月	1.196	1.672	1.350	1.297	1.320	1.278	1.248
MAE		0.595	0.265	0.234	0.248	0.328	0.229
MPAE		0.515	0.229	0.202	0.215	0.284	0.198
P		0.485	0.771	0.798	0.785	0.716	0.802
RNL		0.551	0.673	0.737	0.714	0.620	0.745
MRE		0.493	0.029	0.142	0.101	0.103	0.071

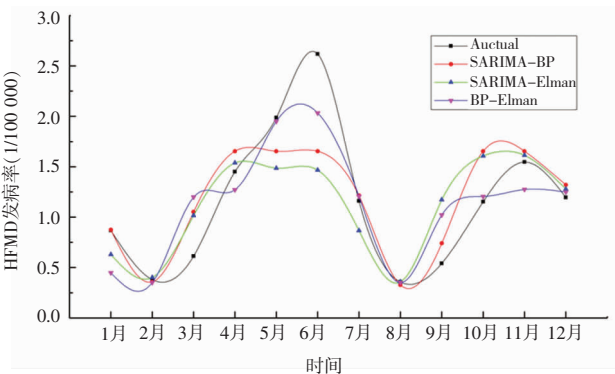


图 3 3 种组合预测模型对 HFMD 发病率的拟合预测曲线图

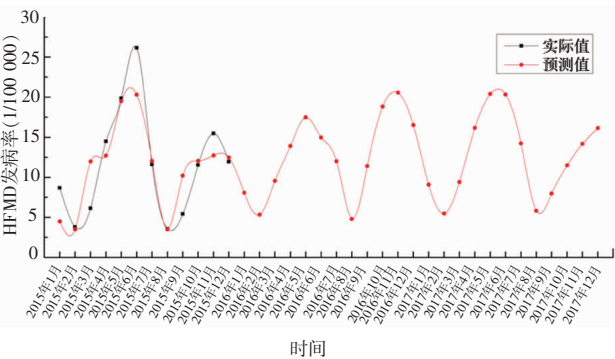


图 4 BPNN-Elman 组合模型对重庆市 2016 至 2017 年 HFMD 发病率预测图

3 讨论

HFMD 传染性强,传播途径复杂,季节性发病高峰明显,是危害我国儿童健康的主要传染病,已对儿童的生命和健康构成严重威胁,并给社会、家庭带

来一定的负担^[13-14]。该病的疫情防控形势严峻,越发地凸显对 HFMD 流行控制的滞后与被动^[15]。在 HFMD 疫情发生和流行之前,准确预测其发生发展和流行趋势,是实现 HFMD 疫情预防和控制的有效途径。因此,探索科学、有效、合理的手足口预警模型具有重大的现实意义。

由于 HFMD 发病率数据一般均存在长期性和周期性变化,ARIMA 模型作为传统的线性模型,能将各种已知的、未知的因素综合统一地蕴含在时间序列变量中,进而实现预测分析。但其无法直观解释时间序列的动态特征及发展趋势,且对时间序列的非线性信息提取效果较差^[16-17]。因此,对于 HFMD 发病率这类具有一定非线性特征的时间序列,其预测的精度往往难以令人满意。而 BP 神经网络模型在处理随机性、非线性数据时有很大的优越性。只要隐含层中节点数足够多,模型就具有逼近任何复杂非线性映射的能力,从而达到对数据波动性的良好预测,被广泛应用于非线性问题的解决^[16-17]。Elman 神经网络是典型的反馈神经网络,它是在 BP 网络的基本结构的基础上,通过存储内部状态使其具备映射动态特征的功能,从而使系统具有适应时变特性的能力。Elman 网络既具有 BP 网络简单明了的优点,又能够更生动、直接地反映系统的动态特性^[18]。但无论是单一的线性模型和非线性模型皆存在信息源不广泛问题,在应用中常常受到限制或达不到高精度的要求^[19],而组合预测模型可以将各种单一的预测模型进行组合,以达到改善预测效果的目的。

这不仅提高了资料的利用率,而且避免了单一模型对数据拟合的局限性。因此,本文建立 3 种单一预测模型(SARIMA、BPNN 和 Elman 神经网络)和 3 种组合预测模型 (SARIMA-BPNN、SARIMA-Elman、BPNN-Elman),实证研究各个模型的预测效能。研究结果表明,预测效果:BPNN-Elman>Elman>SARIMA-BPNN>BPNN>SARIMA-Elman>SARIMA。BPNN-Elman 和 SARIMA-BPNN 组合模型拟合和预测效果均优于单一的模型,提示组合模型可以实现模型间的优势互补,达到提高预测精度和增加预测稳定性的效果。但 SARIMA-Elman 组合模型预测效果较单独建立的 Elman 网络模型差,原因可能是因为 SARIMA 适用于平稳序列,用于具有波动性和周期性的 HFMD 疫情预测时,预测值与实际值的相对误差范围较大,降低了组合模型预测准确率。有文献报道^[29-30],模型拟合及预测的平均相对误差小于 10%时,用于预测时准确度和精度较高。本研究中 BPNN-Elman 组合模型预测性能最优,且平均相对误差为 7.08%,提示 BPNN-Elman 组合模型可以有效预测重庆市 HFMD 发病率。

利用已建立的 BPNN-Elman 模型预测出重庆市 2016 年和 2017 年 HFMD 年均发病率分别为 162.8/10 万、160.9/10 万,均高于 2015 年发病率。提示重庆市应继续加强 HFMD 的防治工作,采取适当的方式预先发布事件威胁的警告,尤其是在 4 至 7 月和 10 至 12 月发病高峰期时段,并采取相应级别的预警行动,有效降低 HFMD 的发病率。

由于本次研究建立的预测模型都从历史疫情数据上反映疾病的发展变化趋势,所以预测结果存在一定的误差。但在实际工作中,HFMD 的发生和传染受到多种综合因素的影响。在今后的传染病预测工作中,若想达到更准确、有效的预测效果,应考虑增加对 HFMD 影响因素数据的搜集,并在此基础上对预测方法作进一步探索和研究。且本次构建的模型只适用于短期预测,若要得到更好的预测效果的模型,应不断更新月发病率,重新拟合新的模型或者对已建立的模型进行回代验证。

综上所述,尽管 BPNN-Elman 组合模型对重庆市 HFMD 发病率预测存在一定误差,但仍可以很好地拟合 HFMD 发病率在时间序列上的变动趋势,对重庆市未来的 HFMD 疫情进行预测,为防控部门有目的、有重点地开展 HFMD 疫情防控工作提供科学依据。

参 考 文 献

- [1] Zhang Y, Tan X, Cui A, et al. Complete genome analysis of the C4 subgenotype strains of enterovirus 71: predominant recombination C4 viruses persistently circulating in china for 14 years[J]. PLoS One, 2013, 8(2): e56341.
- [2] Soto SM. Human migration and infectious diseases[J]. Clin Microbiol Infect, 2009, 15(S1): 26-28.
- [3] 曹 洋, 刘莹莹, 于秋丽, 等. 2010—2015 年河北省手足口病流行特征分析[J]. 疾病监测, 2017, 32(1): 29-33.
- [4] 李法锦, 王 丹, 韦小瑜, 等. 2015 年贵州省手足口病死亡病例流行特征和肠道病毒 71 型 VP1 区基因特征[J]. 中国疫苗和免疫, 2016, 22(6): 611-615.
- [5] 盛振海, 项 东, 刘振荣, 等. 河北省唐山市 2014 年手足口病流行特征及相关影响因素分析[J]. 医学动物防制, 2017, 33(2): 192-194.
- [6] 林 梅, 朱占松, 路 冰. 呼伦贝尔市 2014—2015 年手足口病流行特征分析[J]. 疾病监测与控制, 2017, 11(2): 133-134.
- [7] 张兴裕, 冯海欢, 肖 雄, 等. Elman 神经网络在手足口病发病率预测中的应用初探[J]. 现代预防医学, 2012, 39(9): 2136-2138, 2141.
- [8] 刘 辉, 马殿梅, 刘晓坤, 等. 应用 ARIMA-BPNN 组合模型预测手足口病发病率[J]. 现代预防医学, 2016, 42(16): 2885-2888, 2896.
- [9] 吴文博, 李虹艾, 万鹏程, 等. 利用遗传算法优化的 ARIMA-BP 组合模型预测手足口病发病趋势[J]. 南华大学学报(医学版), 2014, 42(6): 572-576.
- [10] 时照华. 探索 ARIMA 模型在呼吸道传染病疫情预测中的应用[D]. 合肥: 安徽医科大学, 2013.
- [11] 王艳姣, 邓自旺, 王耀庭, 等. 基于 BP 和 Elman 神经网络的福建省汛期旱涝预测模型[J]. 南京气象学院学报, 2004, 27(6): 776-783.
- [12] 雍永强. 基于 ARIMA 和 BPNN 的组合预测模型在血糖预测中的应用[D]. 郑州: 郑州大学, 2015.
- [13] 王连森, 毕振强, 房玉英, 等. 2008 年山东省手足口病流行病学分析[J]. 山东医药, 2009, 49(19): 45-47.
- [14] Zhang Y, Tan XJ, Wang HY, et al. An outbreak of hand, foot, and mouth disease associated with subgenotype C4 of human enterovirus 71 in Shandong, China[J]. J Clin Virol, 2009, 44(4): 262-267.
- [15] 陈澜祺, 王 玲, 吴中发, 等. 手足口病流行的预警方法探讨[J]. 赣南医学院学报, 2010, 30(3): 419-420.
- [16] Khashei M, Bijari M. A new hybrid methodology for nonlinear time series forecasting[M]. Hindawi Publishing Corp, 2011.
- [17] 钟 球, 蒋 莉, 周 琳, 等. 广东省结核病发病趋势的时间序列分析[J]. 中国防痨杂志, 2010, 32(9): 515-519.
- [18] 何 定, 徐 鹏. Elman 与 BP 神经网络应用于交通流预测的对比研究[J]. 工业工程, 2010, 13(6): 97-100.
- [19] 张国良, 后永春, 舒 文, 等. 三种模型在肺结核发病预测中的应用[J]. 中国卫生统计, 2013, 30(4): 480-483.
- [20] 王永斌, 郑 瑶, 柴 峰, 等. 基于周期分解的 ARIMA 模型在甲肝发病率预测中的应用[J]. 现代预防医学, 2015, 42(23): 4225-4229.
- [21] Liu L, Luan RS, Yin F, et al. Predicting the incidence of hand, foot and mouth disease in Sichuan province, China using the ARIMA model—CORRIGENDUM[J]. Epidemiol Infect, 2016, 144(1): 152.

(责任编辑: 冉明会)