

小波去噪后极端学习机模型在医院感染发病率预测中的应用

王清青¹, 范馨月², 查筑红¹

(1. 贵州医科大学附属医院医院感染管理科, 贵州 贵阳 550004;

2. 贵州大学数学与统计学院, 贵州 贵阳 550025)

摘要:目的 采用小波去噪后极端学习机模型预测某三甲医院的医院感染发病率, 并与原极端学习机模型和神经网络模型的预测效果进行比较。**方法** 将2014年1月–2019年6月某三甲医院的医院感染发病率数据作为训练集构建小波去噪后极端学习机模型、原极端学习机模型和神经网络模型, 分别对2019年7月–12月医院感染发病率进行预测。用三类误差分析指标刻画三种模型的拟合效果, 并比较其预测效果。选用较优预测模型预测2020年1月–12月该院感染发病率。**结果** 该院2014–2019年各年医院感染发病率比较, 差异有统计学意义($P<0.05$), 且存在下降趋势($P<0.05$); 小波去噪后极端学习机模型、原极端学习机模型和神经网络模型预测误差评价 MAPE 分别为 0.89%、2.99% 和 5.28%, MRE 分别为 0.01、0.03 和 0.05, RSE 分别为 0.02、0.04、0.05; 选用小波去噪后极端学习机模型对 2020 年 1–12 月该院感染发病率进行预测, 预测值均在 95%CI 范围内。**结论** 小波去噪后极端学习机模型预测医院感染发病率的各项误差指标均低于原极端学习机模型和神经网络模型, 具有较好的拟合与预测效果, 可为预防与控制医院感染的发生提供一定的理论参考。

关键词: 小波去噪后极端学习机; 医院感染; 发病率; 预测模型

中图分类号: R197

文献标识码: A

DOI: 10.3969/j.issn.1006-1959.2022.08.003

文章编号: 1006-1959(2022)08-0008-05

Application of Extreme Learning Machine Model Based on Wavelet Denoising in Prediction of Healthcare-associated Infection

WANG Qing-qing¹, FAN Xin-yue², ZHA Zhu-hong¹

(1. Department of Healthcare-associated Infection Management, the Affiliated Hospital of Guizhou Medical University, Guiyang 550004, Guizhou, China;

2. School of Mathematics and Statistics, Guizhou University, Guiyang 550025, Guizhou, China)

Abstract: **Objective** To study the apply extreme learning machine model based on wavelet denoising in prediction of the healthcare-associated infection incidence, and compare the prediction effect with extreme learning machine model and neural network model. **Methods** The healthcare-associated infection incidence data of a hospital from January 2014 to June 2019 were used as training sets to establish the extreme learning machine model based on wavelet denoising, the extreme learning machine model and the neural network model. To predict the healthcare-associated infection incidence of this hospital from July to December 2019 using the three models, respectively. Three kinds of error analysis indexes were used to describe the fitting effect of the three models, and the prediction effect was compared. The healthcare-associated infection incidence from January to July 2020 was predicted with the best prediction model. **Results** The incidence of healthcare-associated infection in the hospital from 2014 to 2019 had a statistically significant difference ($P<0.05$), and showed a downward trend ($P<0.05$). The MAPE of extreme learning machine model based on wavelet denoising, extreme learning machine model and neural network model was 0.89%, 2.99% and 5.28%, respectively; the MRE was 0.01, 0.03 and 0.05, respectively; the RSE was 0.02, 0.04 and 0.05, respectively. The extreme learning machine model based on wavelet denoising was used to predict the incidence of healthcare-associated infection from January to December 2020, and the predicted values were all within the 95%CI range. **Conclusion** The error indexes of the extreme learning machine based on wavelet denoising are lower than those of the extreme learning machine model and neural network model, which has good fitting and prediction effect, and can provide certain theoretical reference for the prevention and control of nosocomial infection.

Key words: Extreme learning machine model based on wavelet denoising; Healthcare-associated infection; Incidence; Prediction model

近年来,医院感染(healthcare-associated Infection, HAI)给患者和社会造成的严重经济负担已成为医学界急需解决的重大问题。医院感染不仅会提高患者死亡风险,还会延长患者的住院时间、降低病床周转率,严重影响医疗质量^[1-3]。因此,建立预测模型准确预测医院感染发病率,提前采取相应防控措施

施,对降低医院感染发病率具有重要的现实意义。医院感染发病率常用的预测方法包括自回归移动平均滑动模型(auto regressive integrated moving average model, ARIMA)^[4-6]、灰色预测模型(gray forecast model, GM)^[7-9]和神经网络模型(neural network model, NNM)^[10-13]等。ARIMA 模型和 GM 模型使用解析模型对数据进行拟合预测,误差较大。NNM 模型易陷入局部最优解,学习效率和模型精度较低。极端学习机(extreme learning machine, ELM)模型为单隐层前馈神经网络(single-hidden layer feedforward neural networks, SLFNs),具有原理简单、训练速度快的特点^[14],但数据的不平衡分布及其冗余信息产生的噪声以及离群点会对 ELM 的泛化能力产生影响。基于

基金项目:1.国家自然科学基金项目(编号:11961008);2.贵州大学线上线下混合式课程建设项目(编号:XJC202060)

作者简介:王清青(1985.1-),女,贵州镇远县人,博士研究生,副主任医师,主要从事医院感染流行病学研究

通讯作者:范馨月(1984.8-),女,贵州安顺人,博士研究生,副教授,主要从事大数据统计建模、大规模计算研究

此,本研究构建了一种基于小波去噪后的 ELM 模型对医院感染率进行预测分析,以期为医院感染预防与控制提供一定的参考依据。

1 资料与方法

1.1 资料来源 选取贵州省某三甲医院 2014 年 1 月—2019 年 12 月收治的住院时间超过 48 h 的所有住院患者的月度医院感染发病率资料。本研究经过医院伦理委员会审核通过,患者及家属签署知情同意书。

1.2 医院感染诊断标准 参照中华人民共和国卫生部发布施行的《医院感染诊断标准》^[15]。

1.3 方法

1.3.1 小波去噪后极端学习机模型 根据噪声与信号在不同频带上的小波分解系数具有不同强度分布的特点,将各频带上的噪声对应的小波系数去除,保留原始信号的小波分解系数,然后对处理后的系数进行小波重构,得到不含噪声的信号。离散型的小波变换定义为

$$DW_y(m,n) = \alpha^{-\frac{m}{2}} \int_{-\infty}^{+\infty} y(t) \varphi^* \left(\alpha^{-\frac{m}{2}} t - n\tau_0 \right) dt$$

其中, m 为尺度参数, n 为平移常数, φ^* 为复共轭函数, $\varphi(x)$ 为母小波。小波去噪后的序列消除了序列的噪声对预测结果的影响。

不同的样本数据 (X_i, Y_i) , 其中 $X=[X_1, X_2, \dots, X_N]^T \in R^N$, $Y=[Y_1, Y_2, \dots, Y_N]^T \in R^m$ 网络中含有 L 个隐层节点

$$\sum_{i=1}^L \beta_i \cdot f(X_i, a_i, b_i) = Y_i, \quad j=1, \dots, N \quad (1)$$

其中, $\alpha_i=(\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{iN})^T$ 为连接 i 个隐层节点与输入节点的输入权重向量, b_i 为第 i 个隐层结点的偏置。 $\beta_i=(\beta_{i1}, \beta_{i2}, \dots, \beta_{iN})^T$ 为连接第 i 个隐层结点和输出结点的输出权重向量, $f(\cdot)$ 为激活函数。(1)式改写为矩阵形式为:

$$H\beta=Y \quad (2)$$

其中, H 是隐层节点的输出, β 为输出权重, Y 为期望输出。

$$H = \begin{bmatrix} f(X_1, a_1, b_1) & \dots & f(X_1, a_L, b_L) \\ \vdots & \dots & \vdots \\ f(X_N, a_1, b_1) & \dots & f(X_N, a_L, b_L) \end{bmatrix} = \begin{bmatrix} h_1 \\ \vdots \\ h_N \end{bmatrix}$$

$Y=(Y_1, Y_2, \dots, Y_N)^T$, $\beta=(\beta_1, \beta_2, \dots, \beta_N)^T$, ELM 的训练过程等价于求解线性方程(2), 隐含层和输出层之间的连接权重 β 不需要迭代调整, 这也是 ELM 最大的特点。 ELM 输入层和隐含层的连接权重、隐含层的阈值可以随机设定, 且设定完后不用再调整, ELM 结构图见图 1。

1.3.2 参数选择 ELM 模型中的参数有嵌入维数、隐含层神经元数、输入权重 α_i 、神经元偏置 b_i 和输出权重。根据预测误差确定嵌入维数和隐含层神经元

数, 这里取嵌入维数为 2, 隐含层神经元数量 3。输入权重和神经元偏置在 $[-1, 1]$ 内随机取值, 输出权重 β 通过求解线性方程(2)的矩阵 H 的 Moore-Penrose 广义逆得到。以 2014 年 1 月—2019 年 6 月医院感染发病率数据作为训练样本构建小波去噪后极端学习机模型、原极端学习机模型和神经网络模型, 分别对 2019 年 7 月—12 月该医院的医院感染发病率进行预测, 并对其预测效果进行比较。用较优的预测模型对 2020 年 1 月—12 月医院感染发病率进行预测。

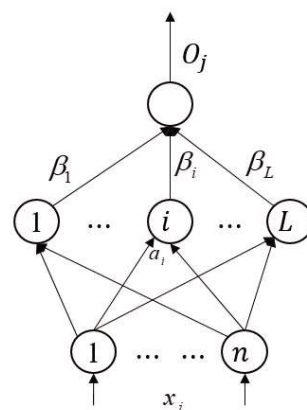


图 1 ELM 结构图

1.4 统计学分析 应用 SPSS 18.0 软件进行统计分析。每年医院感染发病率之间比较采用 χ^2 检验, $P < 0.05$ 表示差异有统计学意义。采用 Matlab 2017a 软件建立小波去噪后极端学习机模型, 模型精度评价采用平均绝对误差百分比 (MAPE)、平均相对误差绝对值 (MRE) 和均方根误差 (RSE)。

1.5 结果分析指标 令 y_i^* 为实际医院感染发病率, \hat{y}_i 为拟合值或预测值。定义三类误差分析指标: 实际值与预测值平均绝对误差百分比 (MAPE) 为:

$$MAPE = \frac{\sum_{i=1}^n \left| \frac{\hat{y}_i - y_i^*}{y_i^*} \right|}{\sum_{i=1}^n y_i^*} \times 100\% \quad (3)$$

相对误差绝对值的平均值 (MRE) 为:

$$MAE = \frac{1}{N} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i^*}{\sum_{i=1}^n y_i^*} \right| \quad (4)$$

均方根误差 (RSE) 为:

$$RSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i^*)^2}{n}} \quad (5)$$

2 结果

2.1 医院感染情况 该院 2014—2019 年医院感染发病率在 0.54%~2.69%, 各年发病率比较, 差异有统计学意义 ($P < 0.05$), 且存在下降趋势 ($P < 0.05$), 见图 2。

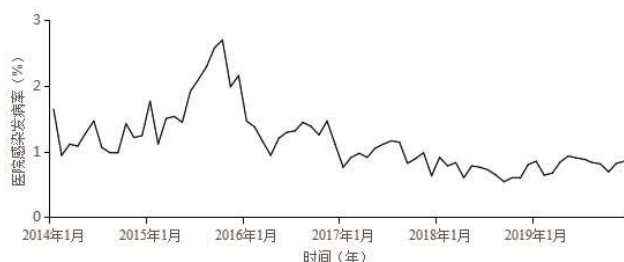


图2 2014-2019年某院医院感染发病率变化趋势

2.2 模型识别 采用小波去噪后 ELM 模型、原 ELM 模型和 NN 模型对 2014 年 1 月-2019 年 6 月该三甲医院医院感染发病率进行训练。小波去噪后 ELM 模型能较好地拟合 2014 年 1 月-2019 年 6 月该三甲医院医院感染发病率,平均绝对误差为 0.11%,均方根误差为 0.01,对原始数据的拟合效果优于原 ELM 模型和 NN 模型,见图 3 及表 1。

2.3 模型诊断 将 2019 年 7 月-12 月医院感染率数据作为验证集。采用本研究建立的小波去噪后极端

学习机模型、原极端学习机模型和神经网络模型对该三甲医院 2019 年 7 月-12 月医院感染率作预测,并和实际值进行对照检验,比较实际值和预测值的误差。小波去噪后极端学习机模型的 MAPE、MRE 和 RSE 均低于原极端学习机模型和神经网络模型,见表 2 及图 4。

2.4 模型预测 选用小波去噪后 ELM 模型对 2020 年 1 月-12 月医院感染发病率进行预测,预测值均在 95%CI 范围内,见图 5。

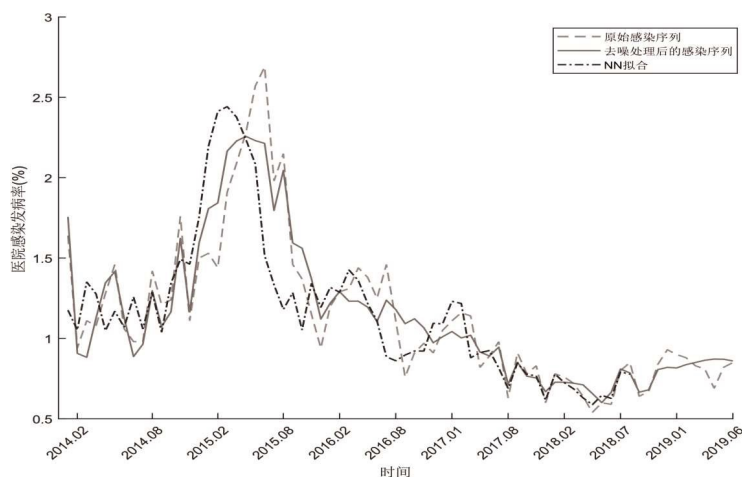


图3 三种预测模型对医院感染发病率的预测拟合效果

表1 三种预测模型对医院感染率拟合误差情况

项目	MAPE(%)	MRE	RSE
NN 模型	70.11	0.55	0.13
原 ELM 模型	1.22	0.33	0.09
小波去噪后 ELM 模型	0.11	0.14	0.01

表2 三种预测模型对医院感染率的预测误差

项目	MAPE(%)	MRE	RSE
NN 模型	5.28	0.05	0.05
原 ELM 模型	2.99	0.03	0.04
小波去噪后的 ELM 模型	0.89	0.01	0.02

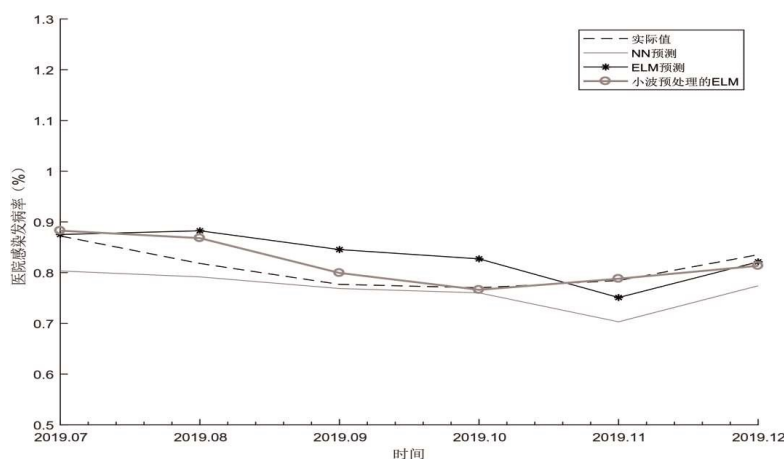


图4 三种预测模型对医院感染发病率的预测效果

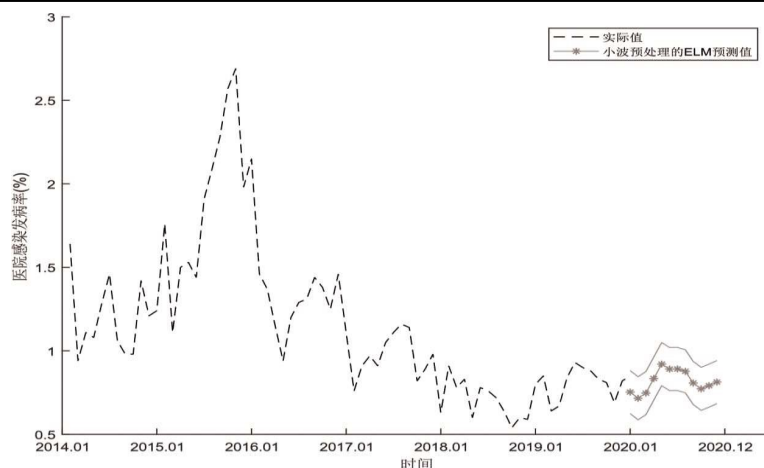


图5 小波去噪后 ELM 模型对医院感染发病率的预测情况

3 讨论

医院感染作为特殊的传染性疾病,其发生、发展与医疗活动息息相关。通过医院感染发病率的前瞻性预测,不仅可以对当前采取的医院感染控制措施效果进行评价,同时还对潜在的院感暴发具有一定的预警作用。因此,基于医院感染监测数据建立准确的医院感染预测、预报体系,使医院管理者从整体上把握医院感染的趋势,从而及时调整相应的预防控制措施,已成为医院感染管理工作的重要发展方向^[16-18]。

医院感染发病率受到季节、环境等因素的影响,数据变化情况较为复杂,如果作为时间序列进行处理难以用解析函数对其进行逼近,导致传统的ARIMA模型和GM(1,1)模型拟合精度较低。对NN模型而言,由于隐函数为非线性函数,其导数值也仅仅在中心值附近呈近似线性性质,中心值的差异性导致线性关系范围的大小不同,使得NN模型难以收敛或仅仅达到局部收敛。ELM只有网络的隐含层节点数需要人为设定,在训练过程中输入权值及神经元偏置随机生成,具有学习和收敛速率快,拟合能力强等优点。因此,ELM算法及其改进算法在医疗诊断^[19]、疾病预测^[20-22]、通信技术^[23]方面都有广泛的应用。本研究构建的小波去噪后ELM模型,有效结合了小波函数的非线性数据处理能力和极端学习机的函数逼近能力。本研究从模型的拟合精度和预测精度进行了对比分析,结果显示,小波去噪后ELM模型的MAPE、MRE和RSE分别为0.89%、0.01、0.02,均小于原极端学习机模型和神经网络模型,具有较好的预测效果。但是,预测的结果仍存在一定误差,可能与样本数据较少、资料所限有关。

医院感染的发生受诸多因素的影响,不同的医院环境其医院感染发病率及变化趋势也不尽相同。因此,医院感染发病率预测模型预测效果的优劣要结合医院实际情况而定,需具体研究。其次,预测模型的构建基于平时的监测数据,后续研究中为进一

步提高模型预测精度,需要不断积累监测数据,对小波去噪后ELM模型参数进行调整,从而修正或重新拟合医院感染发病率预测模型。另外,可以考虑将医院感染相关影响因素纳入模型,以便提高结果的准确性。

参考文献:

- [1]Alp E,Damani N.Healthcare-associated infections in intensive care units: epidemiology and infection control in low-to-middle income countries [J].J Infect Dev Ctries,2015,9 (10):1040-1045.
- [2]Schreiber PW,Sax H,Wolfensberger A,et al.The preventable proportion of healthcare-associated infections 2005-2016: Systematic review and meta-analysis [J].Infect Control Hosp Epidemiol,2018,39(11):1277-1295.
- [3]Cassini A,Plachouras D,Eckmanns T,et al.Burden of Six healthcare-associated infections on european population health: estimating incidence-based disability-adjusted life years through a population prevalence-based modelling study [J].PLoS Med, 2016,13(10):e1002150.
- [4]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.
- [5]周明,李晓玲.住院新生儿医院感染发生季节 ARIMA 模型预测[J].中国消毒学杂志,2019,36(2):107-110.
- [6]Zhao Y,Ge L,Zhou Y,et al.A new Seasonal Difference Space-Time Autoregressive Integrated Moving Average (SD-STARIMA) model and spatiotemporal trend prediction analysis for Hemorrhagic Fever with Renal Syndrome (HFRS)[J].PLoS One,2018,13(11):e0207518.
- [7]Wang YW,Shen ZZ,Jiang Y.Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China[J].PLoS One,2018,13(9):e0201987.
- [8]Yang X,Zou J,Kong D,et al.The analysis of GM (1,1) grey model to predict the incidence trend of typhoid and paratyphoid fevers in Wuhan City, China [J].Medicine (Baltimore),2018,97 (34):e11787.

- [9]Guo X,Liu S,Wu L,et al.Application of a novel grey self-memory coupling model to forecast the incidence rates of two notifiable diseases in China: dysentery and gonorrhea [J].PLoS One,2014,9(12):e115664.
- [10]Akhtar M,Kraemer MUG,Gardner LM.A dynamic neural network model for predicting risk of Zika in real time [J].BMC Med,2019,17(1):171.
- [11]Chen J,Chen J,Ding HY,et al.Use of an Artificial Neural Network to Construct a Model of Predicting Deep Fungal Infection in Lung Cancer Patients [J].Asian Pac J Cancer Prev, 2015,16(12):5095-5099.
- [12]Lammers RL,Hudson DL,Seaman ME.Prediction of traumatic wound infection with a neural network-derived decision model [J].Am J Emerg Med,2003,21(1):1-7.
- [13]Zhu X,Fu B,Yang Y,et al.Attention-based recurrent neural network for influenza epidemic prediction [J].BMC Bioinformatics,2019,20(Suppl 18):575.
- [14]刘长良,吴家佳.基于核极端学习机的短期风电功率预测研究 [J].热能动力工程,2017,32(1):95-100.
- [15]中华人民共和国卫生部.医院感染诊断标准(试行) [J].中华医学杂志,2001,81(5):314-320.
- [16]陈超,张倩.极端学习机模型在衡水市乙肝月发病率预测中的应用 [J].现代预防医学,2018(45):208-212.
- [17]杨旭,张倩.极端学习机模型在张家口市手足口病发病率预测中的应用 [J].第二军医大学学报,2018(39):226-230.
- [18]张倩,陈超,夏晴,等.基于 EMD 分解的极端学习机模型在衡水市乙肝发病率预测中的应用 [J].现代预防医学,2018(45): 1733-1736.
- [19]Ding S,Zhang N,Xu X,et al.Deep extreme learning machine and its application in EEG classification [J].Math Probl Eng,2015 (1):1-11.
- [20]王青青,范馨月,查筑红,等.基于乘积季节性 ARIMA 模型对神经内科医院感染发病率的预测研究 [J].中国感染控制杂志,2019(18):59-63.
- [21]王力红,赵霞,张京利,等.重症监护病房医院感染预警模型的建立 [J].中华医院感染学杂志,2010,20(21):3368-3370.
- [22]张积平,刘学工,刘艳萍,等.医院感染预警控制模型对控制 ICU 医院感染可行性研究 [J].中华医院感染杂志,2015,25(8): 1906-1908.
- [23]Malathi V,Marimuthu NS,Baskar S,et al.Application of extreme learning machine for series compensated transmission line protection [J].Eng Appl Artif Intel,2011(24):880-887.

收稿日期:2021-08-25;修回日期:2021-09-11

编辑/成森