

Elman 神经网络与 ARIMA 模型 对甲肝月发病率的预测研究

安淑一, 方兴, 姚文清

辽宁省疾病预防控制中心, 辽宁 沈阳 110005

摘要: **目的** 两模型对甲肝月发病率数据进行拟合预测, 比较最优模型。 **方法** 通过软件实现 ARIMA 模型和 Elman 神经网络对甲肝发病率进行拟合, 并对 2017 年月发病率进行仿真。 **结果** 两模型拟合预测效果较好, ARIMA 模型平均绝对误差 (mean absolute error, MAE)、均方根误差 (root-mean-square error, RMSE)、平均绝对百分比误差 (mean absolute percentage error, MAPE) 分别为 0.013、0.002 9、9.29; Elman 神经网络 MAE、RMSE、MAPE 分别为 0.012、0.000 22、8.695。Elman 神经网络预测结果优于 ARIMA 模型。 **结论** 两模型均能够拟合预测甲肝月发病率, Elman 神经网络拟合预测效果更好。

关键词: Elman 神经网络; ARIMA; 甲肝; 预测

中图分类号: R512.6⁺1 文献标识码: A 文章编号: 1006-3110(2019)04-0409-04 DOI: 10.3969/j.issn.1006-3110.2019.04.007

Elman neural network and ARIMA model for prediction of monthly incidence of hepatitis A

AN Shu-yi, FANG Xing, YAO Wen-qing

Liaoning Provincial Center for Disease Control and Prevention, Shenyang, Liaoning 110005, China

Corresponding author: YAO Wen-qing, E-mail: wenqingyao@sina.com

Abstract: **Objective** To apply two models to fit the data of the monthly incidence of hepatitis A, to forecast the monthly incidence in 2017, and to select the optimal model. **Methods** The monthly incidence of hepatitis A was fitted by ARIMA model and Elman neural network, and the monthly incidence rates in 2017 were simulated. **Results** It was evident that the fitting precision of the two models was high, and the fitting and forecast effects were satisfactory. Mean absolute error (MAE), root-mean-square error (RMSE) and mean absolute percentage error (MAPE) based on ARIMA model and Elman neural network were 0.013 and 0.012, 0.002,9 and 0.000,22, 9.29 and 8.695 respectively. The prediction results of Elman neural network were superior to those of ARIMA model. **Conclusions** Both the models can be used to predict the monthly incidence rate of hepatitis A, but the fitting and forecast effects of Elman neural network are better.

Key words: Elman neural network; ARIMA; hepatitis A; forecast

传染病的预测预警工作一直是疾病防控工作的重中之重。多年来,在疾病防控工作研究上,国内外有众多的专家学者开展了各种形式的研究^[1-4],从定性到定量,从简单到复杂,预测的疾病种类越来越多,方法不断推陈出新。

ARIMA 模型是处理时间序列常用模型,它依托时间序列各个数据变量之间的相关,通过现有的已知时间序列变量去预测未知时间序列变量。它通过已知时间序列数据之间的相关性,通过数学模型表达变量,代入预测自变量,进而预测因变量。这种模型已经被广

泛应用到金融、工商、化工、建筑等行业,它所需已知变量数据少,适应性广,是应用于平稳时间序列比较广泛的模型。

神经网络是近年来兴起的精度高、泛化性强的一种预测预警方法。本研究采用的 Elman 神经网络有更强的学习能力,已经被广泛应用于交通、环境、经济领域^[5-8],但是在医药领域应用较少。因此本研究利用甲肝月发病率分别采用 ARIMA 方法和 Elman 神经网络两种方法,分别构建预测模型,并比较两种模型预测效果,为疾病预测预警提供研究基础。

基金项目: 国家科技重大专项 (2017zx10103007-001)

作者简介: 安淑一 (1981-), 女, 辽宁大连人, 硕士, 主要从事传染病预防控制工作。方兴 (1978-), 男, 辽宁锦州人, 硕士, 主要从事传染病预防控制工作。二者为并列第一作者。

通信作者: 姚文清, E-mail: wenqingyao@sina.com。

1 资料与方法

1.1 资料来源 资料来自《卫生部公报》和《卫生统计年鉴》。本研究提取 2011 年 1 月-2017 年 12 月甲肝疫情监测资料。

1.2 ARIMA 模型建立

1.2.1 数据预处理 ARIMA 模型使用的前提条件是平稳序列。通过数据变换,使达到平稳时间序列的目的。本研究采用的序列,有季节性成分,因此采用乘积季节模型进行计算。模型初始表示为 $ARIMA(p,d,q)(P,D,Q)_s$ 。其中 d 和 D 为差分和季节性差分次数。其中 p,q,P 和 Q 分别表示连续模型和季节模型中的自回归阶数和移动平均阶数。 s 是季节周期。

1.2.2 模型定阶 通过绘制 ACF 图和 PACF 图,对模型阶数进行判别。

1.2.3 模型检验 建立模型并进行参数估计,并检验模型。

1.2.4 模型预测 成功建立模型后,应用模型进行外推。

1.3 Elman 神经网络建立

1.3.1 数据归一化处理 所有数据归一化处理,所有数据落入 $[0,1]$ 区间。

1.3.2 数据导入 本研究数据为 2011-2017 年月发病率数据,以第 1~12 个数据做为输入组 1,第 13 个数据作为输出组 1,第 2~13 个数据为输入组 2,第 14 个数据为输出组,以此类推。2011-2016 年数据作为神经网络拟合数据集,2017 年数据作为网络预测数据集。

1.3.3 网络的建立和训练 使用 MATLAB 2015b 软件,通过神经网络工具箱建立 Elman 神经网络。针对网络隐含层神经元个数,目前没有好的确定方法,一般不会超过 20。通过误差参数对本研究隐含层最优神经元个数进行择优。

1.4 两种模型的检验与评价 本研究两种模型比较采用如下几个参数^[9-10]。采用平均绝对误差 (mean absolute error, MAE) 为预测误差绝对值的平均值,均方根误差 (root-mean-square error, RMSE) 是观测值与真值偏差的平方和观测次数 n 比值的平方根,平均绝对百分比误差 (mean absolute percentage error, MAPE),预测误差绝对值与真值比值求和后与观测次数 n 的比值的百分数。这 3 个指标评价模型的预测效能,MAE、RMSE、MAPE 这三个参数的值越小,表明模型的预测效果越好。具体公式如下:

$$MAE = \frac{\sum_{i=1}^n |e_i|}{n}, RMSE = \frac{\sum_{i=1}^n e^2}{n}, MAPE = \frac{\sum_{i=1}^n \frac{|e_i|}{\chi_i}}{n} * 100$$

1.5 统计方法 利用 Excel 2010 对收集到的数据进行初步整理。利用 SPSS 22.0 和 MATLAB2015b 软件进行两种模型构建,统计运算,处理和分析。

2 结果

2.1 ARIMA 模型建立 (1)平稳序列变换。数据通过一次季节差分后,满足平稳序列要求。 D 为 1。(2)通过 ACF 图(见图 1)和 PACF 图(见图 1、图 2)初步判断 p 为 1, q 为 1。根据文献 P,Q 取值一般不会大于 2。通过 AIC、BIC 辅助判断,最终确定最优 $P=1,Q=1$ 。(3)确定最终最优模型形式为 $ARIMA(1,0,1)(1,1,1)_{12}$,模型拟合参数最好,AIC=2.425,BIC=7.832,模型残差的自相关和偏自相关系数均落在 2 倍标准差之内,残差序列呈现白噪声,见图 3。 $ARIMA(1,0,1)(1,1,1)_{12}$ 型,预测结果均方根误差 (RMSE) 为 0.0029,平均绝对误差 (MAE) 为 0.013,平均绝对百分比误差 (MAPE) 为 9.29,表明模型预测结果较为理想。(4)通过模型预测 2017 年数据。

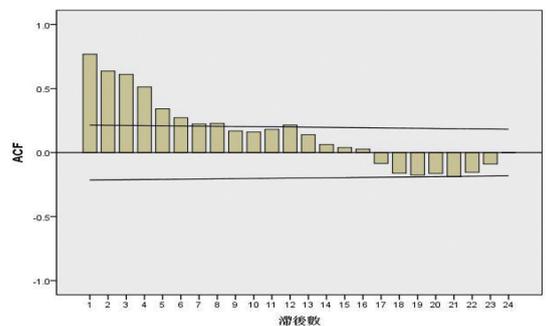


图 1 ARIMA 模型 ACF 图

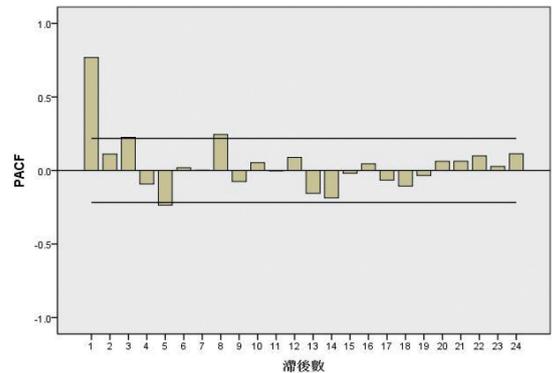


图 2 ARIMA 模型 PACF 图

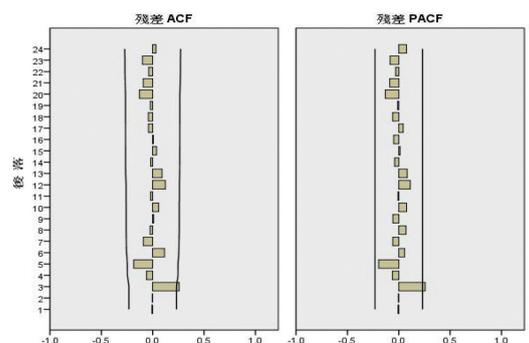


图 3 SARIMA 模型残差的自相关和偏自相关

2.2 Elman 神经网络建立

2.2.1 网络训练 本研究输入层节点为 12, 输出层节点为 1。共 7 年的数据, 前 6 年的数据作为训练, 最后 1 年数据作为预测数据。以误差 0.000 1 为训练目标, 1 000 为训练步长误差快速收敛。

2.2.2 最优隐含层节点数探讨 隐含层神经元的个数是比较复杂的问题, 隐含层神经元数目的选择目前还没有确切的理论和方法, 本研究通过逐渐增加和减少隐含层神经元数目的方法, 选择使 MAE 最小的隐含层神经元数目, 最后确定最优隐含层神经元为 9, 测试表明神经元的个数再增加, MAE 逐渐增大。见表 1。

表 1 不同隐含层神经元 MAE 情况

编号	隐含层神经元	MAE	编号	隐含层神经元	MAE
1	1	0.072 9	11	11	0.057 3
2	2	0.059 4	12	12	0.053 2
3	3	0.047 6	13	13	0.057 7
4	4	0.064	14	14	0.054 7
5	5	0.079 4	15	15	0.060 6
6	6	0.066 7	16	16	0.052 1
7	7	0.063 9	17	17	0.056 9
8	8	0.058 6	18	18	0.050 5
9	9	0.041 8	19	19	0.055 8
10	10	0.056 5	20	20	0.057 7

2.2.3 Elman 神经网络 经过 1 000 次训练后, 其训练误差为 0.000 16。网络训练误差迭代收敛曲线见图 4。

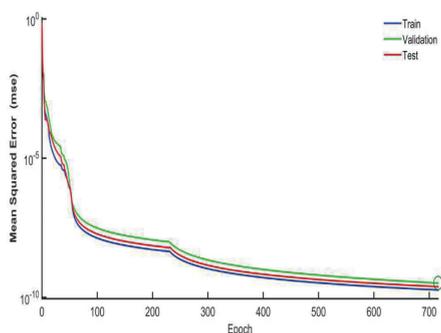


图 4 Elman 神经网络误差迭代收敛曲线

2.2.4 预测误差 Elman 神经网络 均方根误差 (RMSE) 为 0.000 22, 平均绝对误差 (MAE) 为 0.012, 平均绝对百分比误差 (MAPE) 为 8.695。ARIMA 和 Elman 神经网络两种模型预测误差均为早期数据误差较大, 随着时间推移误差逐渐减小, 见图 5。通过 ARIMA 和 Elman 神经网络分别对 2017 年发病率进行预测比较, 见表 2。

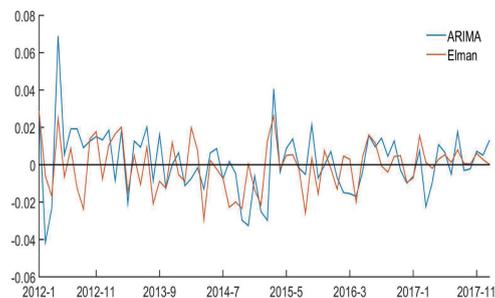


图 5 Elman 神经网络和 ARIMA 模型拟合误差曲线

表 2 两模型 2017 年预测结果比较

日期(年-月)	实际值(1/10 万)	ARIMA 预测发病率(1/10 万)	Elman 预测发病率(1/10 万)
17-1	0.103 3	0.110 8	0.118 7
17-2	0.116	0.093 6	0.117 3
17-3	0.142	0.132 5	0.140 1
17-4	0.120 9	0.131 7	0.123 8
17-5	0.115 3	0.121 9	0.120 7
17-6	0.115 5	0.110 4	0.116 7
17-7	0.111 7	0.129 4	0.119 7
17-8	0.127 4	0.124 3	0.128 1
17-9	0.128 3	0.126 1	0.128 7
17-10	0.109 9	0.117 1	0.115 9
17-11	0.114 7	0.119 8	0.117 4
17-12	0.102 5	0.115 4	0.102 8

3 讨论

ARIMA 模型作为经典时间序列模型的一种, 被广泛应用到科技、工程、建筑、股票等各行各业, 作为预测模型的一种近年来在医疗行业也广为流行。它可以通过较少的样本数据集, 动态的预测序列变化发展趋势, 它能够把样本数据的波动, 随时反应出来, 其用时短、预测精度好, 被研究者广泛应用。但是随着数据的增多, 波动幅度的变异程度变大, 它在应对一些非线性样本处理, 及不规则数据的处理, 难度增大。对于甲肝类似疾病发病规则性不明显的前提下, 预测精度有所下降。Elman 神经网络是动态神经网络, 其独特的反馈神经元, 保证网络实时学习, 实时更新, 在样本数据集变化较大, 样本数据不需要很多的情况下, 往往能有更好的拟合精度。Elman 神经网络的特点是隐含层的输出通过承接层的延迟与存储, 自联到隐含层的输入, 这种自联方式使其对历史状态的数据具有敏感性, 内部反馈网络的加入增加了网络本身处理动态信息的能力, 从而达到动态建模的目的^[11]。

本研究建立的甲肝发病率预测模型, 采用 Elman 神经网络预测结果各项误差指标 MAE、MAPE 以及 RMSE 结果均比 ARIMA 模型要小。表明 Elman 神经

人参皂苷 Rb1 对周围神经损伤后大鼠背根神经节神经生长因子及其受体酪氨酸激酶 A 表达的影响

邓明¹, 谢萍², 马永刚¹, 刘志勇¹, 周炎¹, 陈庆¹, 刘世清¹, 吴飞¹, 明江华¹

1. 武汉大学人民医院, 湖北 武汉 430060; 2. 武汉市第三医院, 湖北 武汉 430060

摘要: 目的 观察人参皂苷 Rb1 对周围神经损伤后大鼠背根神经节神经生长因子(nerve growth factor, NGF)及其受体酪氨酸激酶 A (tyrosine kinase A, TrkA) 表达的影响。方法 将 SD 大鼠随机分为 5 组:假手术组、模型组、甲钴胺组(10.8 mg/kg)、人参皂苷 Rb1 低剂量组(25 mg/kg)、人参皂苷 Rb1 高剂量组(50 mg/kg)。采用神经夹持损伤法建立周围神经损伤模型,造模完成后灌胃给药,持续 14 d。末次给药后,采用光热耐痛阈测定法评价大鼠的感觉恢复情况,Western-blotting 法检测大鼠背根神经节 NGF 及其受体 TrkA 蛋白表达情况,RT-PCR 法检测基因表达情况。结果 与假手术组比较,模型组大鼠对热痛感觉的灵敏度下降(11.23±2.11 vs. 7.14±0.67, $P < 0.01$),背根神经节 NGF 及其受体 TrkA 蛋白和基因表达均显著下调(0.34±0.05 vs. 0.87±0.11, 0.31±0.04 vs. 0.79±0.10, $P < 0.01$);给予高剂量的人参皂苷 Rb1 干预后,模型大鼠的热痛感觉恢复良好,数值为(8.36±0.92) ($P < 0.01$),背根神经节 NGF、TrkA 的蛋白和基因水平均得到回升,分别为(0.65±0.07)、(0.70±0.08) ($P < 0.01$)。结论 人参皂苷 Rb1 能通过提高背根神经节 NGF 及其受体 TrkA 的表达,维持神经元存活,促进受损神经修复,改善周围神经损伤后大鼠的感觉功能。

关键词: 周围神经损伤; 人参皂苷 Rb1; 背根神经节; 神经生长因子; 酪氨酸激酶 A

中图分类号: R-332 **文献标识码:** A **文章编号:** 1006-3110(2019)04-0412-04 **DOI:** 10.3969/j.issn.1006-3110.2019.04.008

Effects of ginsenoside Rb1 on the expression of nerve growth factor and its receptor tyrosine kinase A in rat dorsal root ganglion after peripheral nerve injury

DENG Ming¹, XIE Ping², MA Yong-gang¹, LIU Zhi-yong¹, ZHOU Yan¹, CHEN Qing¹, LIU Shi-qing¹, WU Fei¹, MING Jiang-hua¹

1. The People's Hospital of Wuhan University, Wuhan, Hubei 430060, China;

2. The Third Hospital of Wuhan City, Wuhan, Hubei 430060, China

Corresponding author: MING Jiang-hua, E-mail: mingjianghua2002@medmail.com.cn

基金项目: 国家自然科学基金青年基金项目(81301052); 中央高校基本科研业务费专项基金(2042017kf0139)

作者简介: 邓明(1983-),男,博士,主治医师,研究方向:神经再生。

通信作者: 明江华, E-mail: mingjianghua2002@medmail.com.cn。

网络预测结果要优于 ARIMA 模型结果。也证明了神经网络在非线性数据及波动不规则数据上效果要优于 ARIMA 模型。建立可靠的信度和精度较高的预测预警模型是当前疾病防控热点之一,有效可靠的预测预警能够带来巨大的社会经济利益,保障人民的健康,通过对各种模型的尝试,能够在疾病暴发流行前,达到预测预警的目的。

参考文献

- [1] Xie JY, Li HL, Ma ZY. Analysis of key factors in heat demand prediction with neural networks[J]. Energy Pro, 2017, 5(105): 2965-2970.
- [2] Imrul Hasan AHMI, Haddawy P, Lawpoolsri S. A comparative analysis of bayesian network and ARIMA approaches to malaria outbreak prediction[J]. ICCI, 2017, 12(3): 108-117.
- [3] 李晓蓉, 庞学文, 于燕明, 等. ARIMA 模型在天津市结核病发病预测中的应用[J]. 实用预防医学, 2018, 25(12): 1536-1538.
- [4] Liang Y, Qiu L, Zhu J, et al. Digester temperature prediction model

based on the Elman neural network[J]. Appl Eng Agric, 2017, 33(2): 143-148.

- [5] 倪飞翔. 基于改进小波-Elman 神经网络算法的蜂窝网流量预测[J]. 电子设计工程, 2017, 25(3): 171-175.
- [6] 杨柳, 吴廷琳, 张超. 改进最小二乘支持向量机电量预测算法[J]. 电网与清洁能源, 2017, 33(5): 71-76.
- [7] 解小平. 基于 Elman 神经网络的短时交通流预测及应用研究[M]. 兰州交通大学学报, 2017, 59(8): 32-40.
- [8] 叶瑞丽, 郭志忠, 刘瑞叶. 基于小波包分解和改进 Elman 神经网络的风电场风速和风电功率预测[J]. 电工技术学报, 2017, 32(1): 34-42.
- [9] 马洁, 田野, 黄璐, 等. ARIMA 乘积季节模型与广义回归神经网络模型在布鲁氏菌病发病预测的比较[J]. 山东大学学报(医学版), 2018, 56(1): 1-6.
- [10] 陈小玲. 基于 ARIMA 模型与神经网络模型的股价预测[J]. 经济数学, 2017, 34(1): 30-34.
- [11] 王阳, 姜燕宁. 基于遗传算法的灰色 Elman 神经网络预测模型及其应用[J]. 数学的实践与认识, 2017, 47(22): 315-320.

收稿日期: 2018-07-09