

· 论 著 ·

应用 Elman 神经网络建立流感样病例预测模型

章涛¹, 官海滨², 李傅冬¹, 何凡¹

1.浙江省疾病预防控制中心公共卫生监测与业务指导所, 浙江 杭州 310051; 2.温州医科大学

摘要: **目的** 应用 Elman 神经网络构建流行性感冒(流感)样病例(ILI)预测模型, 为浙江省流感疫情早期预警提供依据。**方法** 收集 2013—2014 年浙江省 11 家流感监测哨点医院的 ILI 报告、ILI 病毒核酸检测结果、气象和空气污染物等资料, 通过时滞相关性分析筛选纳入模型的变量, 采用 2013 年第 14 周—2014 年第 44 周的数据建立 Elman 神经网络预测模型, 采用 2014 年第 45—52 周的数据检验模型的预测效能。**结果** 浙江省 2013—2014 年每周均有 ILI 报告, 共报告 506 391 例次, 周报告 ILI% 为 $(3.07 \pm 0.73) \%$ 。筛选出提前 13 周的周平均气压、提前 11 周的周平均水汽压、提前 9 周的周平均气温、提前 5 周的周平均 SO_2 浓度、提前 5 周的周平均 NO_2 浓度、提前 5 周的周平均 CO 浓度、提前 5 周的周平均 $\text{PM}_{2.5}$ 浓度、提前 5 周的周平均 PM_{10} 浓度、提前 5 周的周平均 AQI 和提前 1 周的病原阳性率 10 个因素纳入模型。当网络结构为 10-15-1-1 时, 构建的 Elman 模型为最优预测模型, 预测结果的平均误差绝对率为 10.58%, 非线性相关系数为 0.876 7。**结论** 利用气象、空气污染指标和流感病原性监测资料建立的 Elman 神经网络 ILI 预测模型预测效果较好, 适用于浙江省流感疫情短期预测。

关键词: Elman 神经网络; 流感样病例; 预测; 气象; $\text{PM}_{2.5}$

中图分类号: R511.7 文献标识码: A 文章编号: 2096-5087 (2019) 02-0113-06

Modeling of influenza-like illness prediction based on Elman neural network

ZHANG Tao*, GUAN Hai-bin, LI Fu-dong, HE Fan

*Department of Public Health Monitoring, Zhejiang Provincial Center for Diseases Control and Prevention, Hangzhou, Zhejiang 310051, China

Abstract: Objective To build a model for influenza-like illness (ILI) prediction based on Elman neural network and to provide evidence for early warning of influenza epidemic in Zhejiang Province. **Methods** The data of ILI from 11 sentinel hospitals, influenza pathogen detection, meteorological factors and air pollutants in Zhejiang Province from 2013 to 2014 were collected. Time-delay correlation analysis was conducted to select variables for modeling. Based on Elman neural network, data from the 14th week of 2013 to the 44th week of 2014 were used as a training set to establish the model and the data from 45th week to 52nd weeks of 2014 were used as a test set for the model performance. **Results** There were ILI reported every week during 2013 and 2014, with a total of 506 391. The percentage of ILI cases per week was $(3.07 \pm 0.73) \%$. Ten variables selected by time-delay correlation analysis were the weekly average values of atmospheric pressure (13 weeks in advance), vapor pressure (11 weeks in advance), temperature (9 weeks in advance), SO_2 (5 weeks in advance), NO_2 (5 weeks in advance), CO (5 weeks in advance), $\text{PM}_{2.5}$ (5 weeks in advance), PM_{10} (5 weeks in advance), air quality index (5 weeks in advance) and positive rate of pathogen (1 weeks in advance). Elman neural network (10-15-1-1) was selected as the optimal model, and the prediction performed well, with 10.58% as the mean error rate and 0.876 7 as the nonlinear correlation coefficient. **Conclusion** This study demonstrated that Elman neural network including variables of meteorological factors, air pollutants and the positive rate of pathogen performed well on the short-term prediction of ILI incidence.

Key words: Elman neural network; Influenza-like illness; Prediction; Meteorological factors; $\text{PM}_{2.5}$

DOI: 10.19485/j.cnki.issn2096-5087.2019.02.002

基金项目: 2015 年浙江省医药卫生平台计划 (2015RCB011); 2016

年浙江省疾病预防控制中心科技英才孵育项目

作者简介: 章涛, 硕士, 医师, 主要从事公共卫生监测管理工作

通信作者: 何凡, E-mail: fhe@cdc.zj.cn

流行性感（流感）是一种由流感病毒引起的急性发热性呼吸道传染病，具有潜伏期短、传播速度快、波及范围广、抗原易变性、人群普遍易感和控制难度大等典型特征，容易引起暴发或流行。近年研究表明，以流感样病例（influenza-like illness, ILI）占门急诊病例总数的比例（ILI%）进行流感预警的灵敏度、特异度和一致性与病毒阳性率指标的接近，且一般可提前 48 h 左右^[1-2]。多种数学模型可用于流感预测^[3-4]，但流感的发生除季节因素作用外，还受到很多其他因素影响，变化趋势复杂，若仅根据历史发病数据或常规的统计学预测模型，则难以保证预测的准确性。神经网络具有非线性映射、自适应学习和鲁棒容错性等优良特性，能较为精确地描述非线性的动态过程^[5]。Elman 神经网络是 Elman 于 1990 年提出的，结构相对简单，运算量小且适合实时系统辨识的动态递归网络，已被广泛应用于信号监测、交通流量预测等领域^[6-7]。该模型在前馈式网络的隐含层中增加了一个承接层作为延时算子，以达到记忆的目的，从而具有适应时变特性的能力，更生动、直接地反映系统的动态特性^[8-9]。本研究收集 2013—2014 年浙江省流感监测哨点医院门急诊就诊数据、气象条件、空气质量和病原检测情况等，分析这些指标与流感样病例发生的相关性，并基于 Elman 神经网络建立浙江省流感样病例预警模型，旨在提早掌握流感的时间变化特征，有效指导流感防控，降低流感造成的社会危害和经济损失。

1 资料与方法

1.1 资料来源 流感样病例资料来源于中国流感监测信息系统中浙江省 11 家流感监测哨点医院的网络直报资料，病原监测资料来源于浙江省各市疾病预防控制中心对流感监测哨点医院流感样病例采样的核酸检测结果，气象资料来源于中国气象科学数据共享服务网（<http://cdc.nmic.cn/home.do>），空气质量资料来源于浙江省环境保护局。所有资料的时间范围为 2012 年 12 月 31 日—2014 年 12 月 28 日，即 2013 年第 1 周至 2014 年第 52 周，共 104 周。根据《全国流感监测方案（2010 年版）》，ILI 定义为体温 ≥ 38 °C，伴有咳嗽或咽痛，缺乏其他实验室诊断依据的可疑病例^[10]。

1.2 方法 Elman 神经网络一般分为输入层、隐含层、承接层和输出层 4 层。输入层、隐含层和输出层的连接类似于前馈式网络，输入层单元仅起信号传输作用，隐含层单元的传递函数可采用线性或非线性函

数，输出层单元起线性加权作用；承接层用于记忆隐含层单元前一时刻的输出值并返回网络输入，可为一步延时算子。

1.2.1 选择输入量 将气象条件和空气质量数据中的日数据预处理为周数据，考虑潜伏期和疾病后效表现期，先对浙江省周 ILI%、气象条件、空气质量和流感病毒阳性率进行时滞相关性分析，筛选出有统计学意义的带延迟数目的影响因素纳入 Elman 神经网络模型。

1.2.2 数据归一化 为防止大数将小数的信息湮没，必须对筛选出的数据进行预处理^[11]。最常用的数据处理方法为归一化处理^[12]，即用某向量除以其长度，使网络输入的数值都在 0 和 1 之间。本研究采用 Min-Max 标准化进行归一化。

1.2.3 设计神经网络结构 选取 2013 年第 14 周至 2014 年第 44 周的数据作为训练集，用于 Elman 神经网络模型的建立。在训练过程中，不断修正神经网络权值和阈值的规则称为训练算法。本研究采用 Levenberg-Marquardt (LM) 算法优化标准 Elman 神经网络^[13]，加快收敛速度。本研究采用试凑法，根据纳入因素的个数选择隐藏层节点数，训练样本数据集，以误差 0.000 1 作为训练目标，训练 1 000 次，选取训练样本拟合结果均方误差（mean square error, MSE）最小的网络作为最优网络。隐含层神经元传递函数选择正切 S 型函数，输出层选择线性传递函数，训练函数选择 LM 算法训练函数。

1.2.4 模型检验与评价 选取 2014 年第 45 周至第 52 周的数据作为检验集，用于检验与评价模型的预测效能。采用误差绝对值（mean absolute error, MAE）、平均误差绝对率（mean error rate, MER）和非线性相关系数（nonlinear correlation coefficient, RNL）评价该模型的预测效能^[8-9]；MER 越小，且 RNL 越接近 1，说明模型拟合（或预测）效果越好。

1.3 统计分析 采用 SPSS 22.0 软件建立数据库，ILI%、气象因素、空气质量因子和流感病毒核酸阳性率资料均不服从正态性、独立性，且散点图未显示明显的非线性关系，故时滞相关分析采用 Spearman 相关分析；采用 MATLAB 2012b 软件的神经网络工具箱建立、检验和评价 Elman 神经网络模型。检验水准 $\alpha=0.05$ 。

2 结果

2.1 基本情况

2.1.1 ILI 报告情况 浙江省 2013—2014 年每周均有

ILI 报告，共报告 506 391 例次，周报告 ILI% 为 (3.07 ± 0.73)%。ILI% 高峰出现在冬季、春季、夏季以及季节交替时。见图 1。

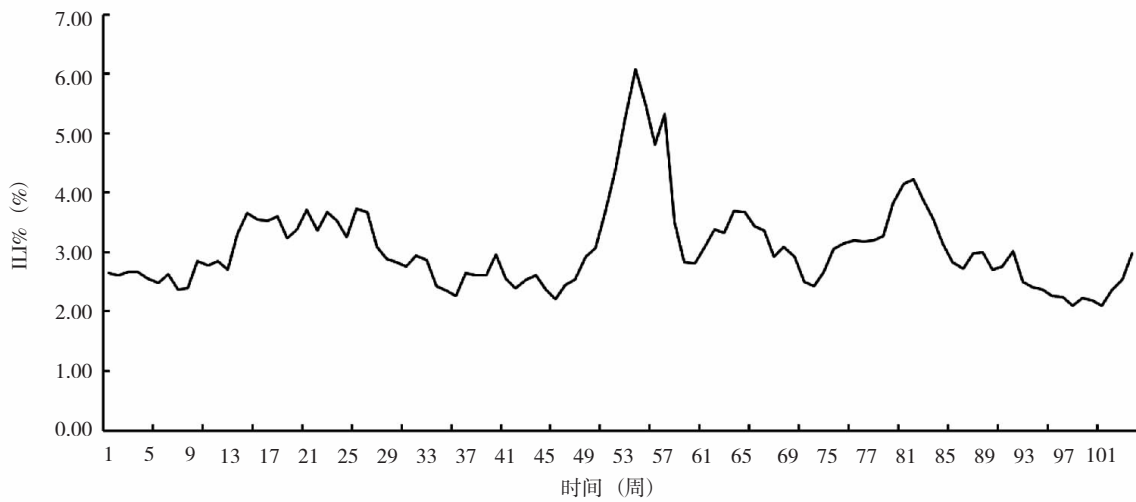


图 1 2013—2014 年浙江省流感监测哨点医院 ILI%

2.1.2 ILI 病原阳性率 2013—2014 年共采集 ILI 标本 30 954 份，检出流感病毒核酸阳性 5 145 份，阳性率为 16.62%；其中甲型流感病毒阳性 3 979 份，占 77.34%；乙型流感病毒阳性 1 158 份，占 22.51%。

2.1.3 气象和空气质量 浙江省 2013—2014 年每周

平均气压、平均 SO₂ 浓度、平均 NO₂ 浓度、平均 CO 浓度、平均 PM_{2.5} 浓度、平均 PM₁₀ 浓度和平均 AQI 呈现夏秋季低、冬春季高的特征；平均水汽压、平均气温呈现夏秋季高、冬春季低的特征。见图 2 和图 3。

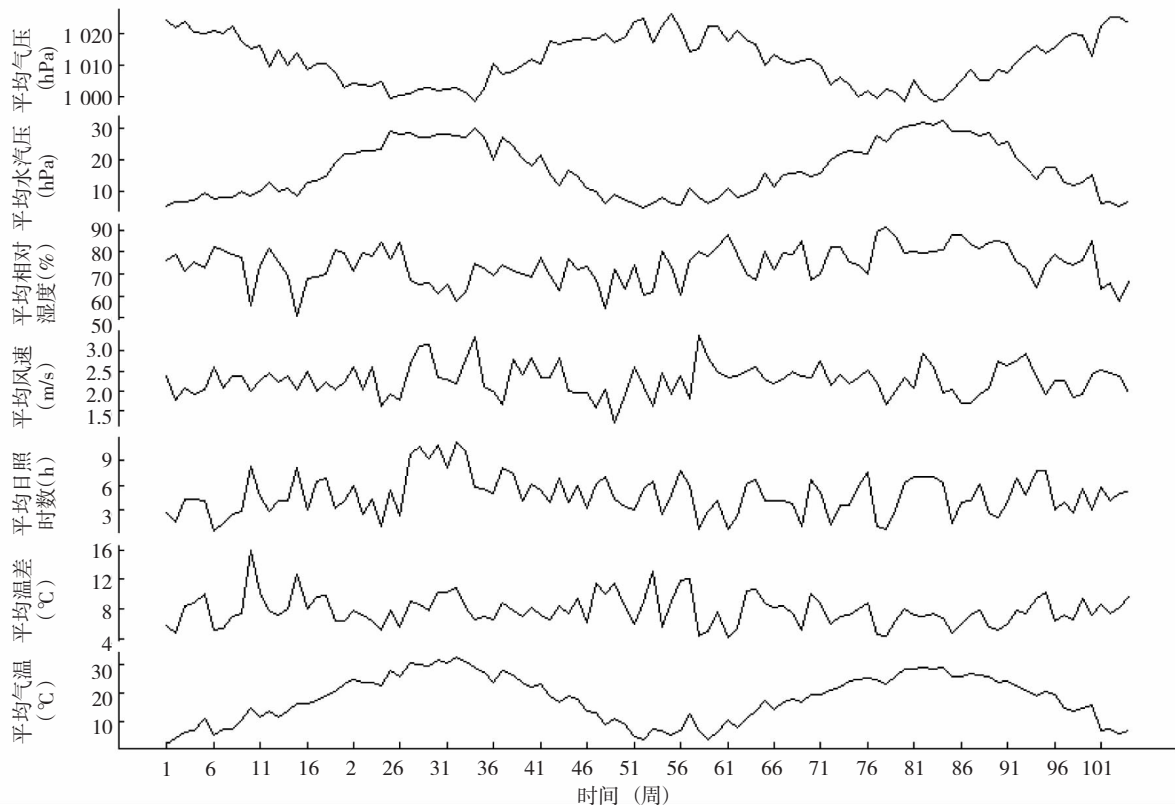


图 2 2013—2014 年浙江省气象指标

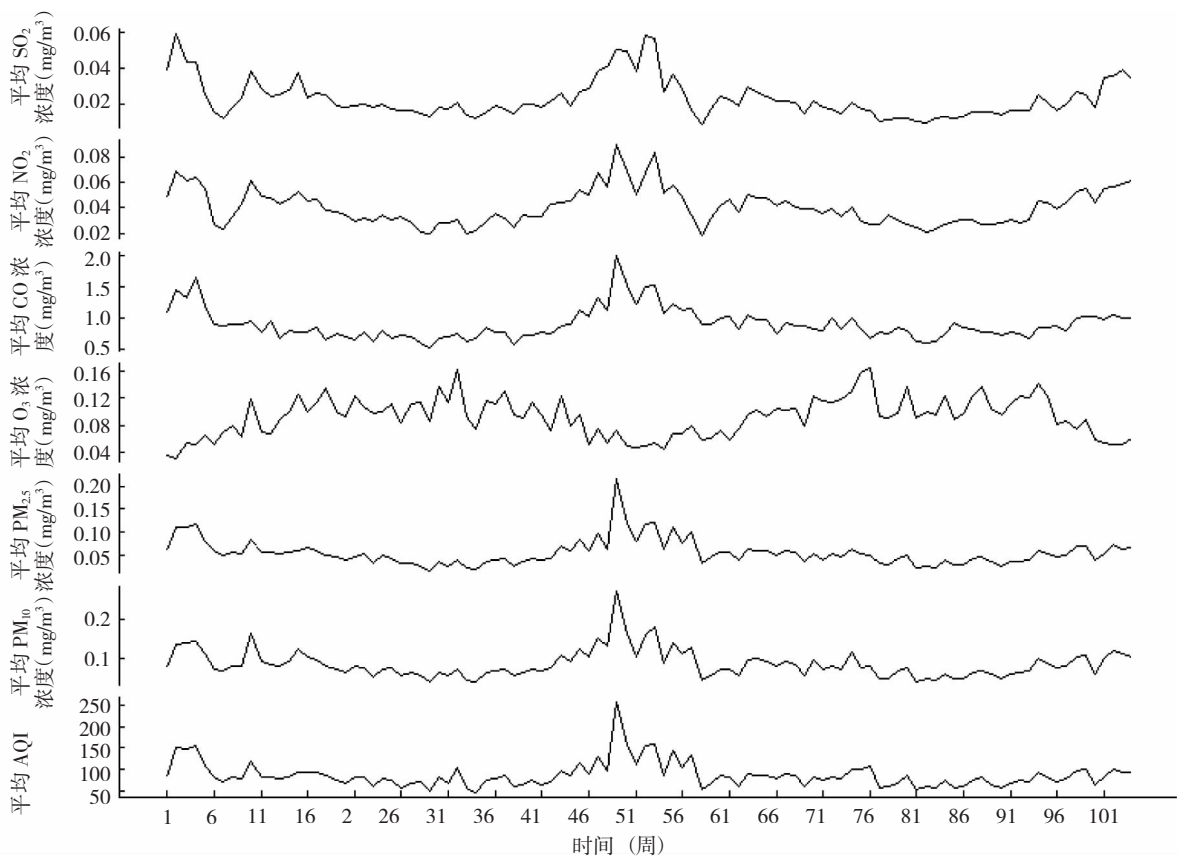


图3 2013—2014年浙江省空气质量指标

2.2 ILI%时滞相关分析 流感病毒核酸阳性率与当周ILI%的相关程度最高 ($r_s=0.619, P<0.001$)；周平均气压与滞后13周的ILI%相关程度最高 ($r_s=0.404, P<0.001$)；周平均水汽压与滞后11周的ILI%相关程度最高 ($r_s=-0.435, P<0.001$)；周平均相对湿度与滞后11周的ILI%相关程度最高 ($r_s=-0.233, P=0.018$)；周平均风速与滞后8周的ILI%相关程度最高 ($r_s=-0.302, P=0.002$)；周平均日照时数与ILI%的相关性无统计学意义 ($r_s=-0.172, P=0.079$)；周平均日温差与滞后10周的ILI%相关程度最高 ($r_s=0.213, P=0.030$)；周平均气温与滞后9周的ILI%相关程度最高 ($r_s=-0.339, P=0.001$)；周平均SO₂浓度与滞后5周的ILI%相关程度最高 ($r_s=0.425, P<0.001$)；周平均NO₂浓度与滞后5周的ILI%相关程度最高 ($r_s=0.416, P<0.001$)；周平均CO浓度与滞后5周的ILI%相关程度最高 ($r_s=0.463, P<0.001$)；周平均O₃浓度与滞后3周的ILI%相关程度最高 ($r_s=-0.251, P=0.010$)；周平均PM_{2.5}浓度与滞后5周的ILI%相关程度最高 ($r_s=0.497, P<0.001$)；周平均PM₁₀浓度与滞后5周的ILI%相关程度最高 ($r_s=0.510, P<0.001$)；周平均AQI与滞后5周后的ILI%的相关程度最高 ($r_s=0.488, P<0.001$)。

2.3 Elman神经网络模型建立 根据上述相关分析结果，同时考虑预测的提前量，最终选取周平均气压（提前13周）、水汽压（提前11周）、气温（提前9周）、SO₂浓度、NO₂浓度、CO浓度、PM_{2.5}浓度、PM₁₀浓度和AQI（均提前5周）以及流感病毒核酸阳性率（提前1周）10个因素进入模型。选取10个神经元作为Elman神经网络的输入层节点数，选取1个神经元作为输出层节点数。以11~24逐个作为隐藏层的节点数分别训练样本数据集，11~24对应的训练样本拟合结果的MSE依次为0.00479、0.00564、0.00382、0.00291、0.00200、0.00389、0.00382、0.00277、0.00280和0.00233，其中以隐含层节点数为15的MSE最小，模型复杂性相对较低，故最优的Elman神经网络模型的结构为10-15-1-1，训练效果见图4。

2.4 模型验证与评价 采用10-15-1-1结构的Elman神经网络模型对浙江省2014年第45—52周的ILI%进行预测，预测结果的MAE为0.25，MER为10.58%，RNL为0.8767，提示该模型预测效能良好。从开始预测的第6周（即2014年第50周）起，预测结果的MER似有上升趋势。见表1。

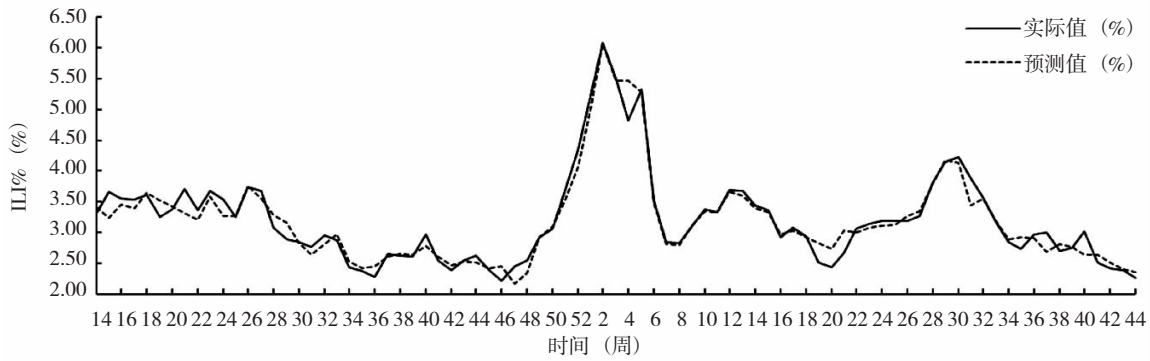


图4 10-15-1-1 结构 Elman 神经网络预测模型的训练效果

表1 浙江省 2014 年第 45—52 周 ILI%10-15-1-1 结构的 Elman 神经网络预测结果(%)

周次	ILI%		MAE	MER
	实际值	预测值		
45	2.25	2.49	0.24	10.67
46	2.10	2.31	0.21	10.00
47	2.24	2.02	0.22	9.82
48	2.20	2.04	0.16	7.27
49	2.10	2.33	0.23	10.95
50	2.36	2.01	0.35	14.83
51	2.54	2.49	0.05	1.97
52	2.98	2.41	0.57	19.13
平均值	2.35	2.26	0.25	10.58

3 讨论

疾病流行动态估计和预测是传染病防制的重要环节，是卫生管理部门制定防控对策、配置资源的主要依据。然而，由于流感的流行受到多种不确定因素的影响，早期识别困难，导致相应的预防控制措施较为滞后，如何及时有效地预警流感流行一直是疾病防制工作的重点和难点^[14]。2013—2014 年浙江省流感监测数据显示，流感活动高峰在冬春季、夏季以及季节交替时，这与我国南方地区的流感活动情况^[15-16]一致。本研究基于浙江省 2013 年第 14 周至 2014 年第 44 周流感疫情资料建立了 10-15-1-1 结构的 Elman 神经网络模型，预测 2014 年第 45—52 周的 ILI%，结果显示，模型预测值与实际监测值接近，变化趋势大致相同，预测效果良好。

孙锦峰等^[9]比较 Elman 神经网络模型与时间序列中常用的 ARIMA 模型，发现 Elman 神经网络具有更好的非线性拟合能力，且该模型在统计学分析软件 MATLAB 中已有工具箱可进行建模和分析，简便、易操作。模型选择是建模的关键技术之一，本研究选

择网络结构为 10-15-1-1 的 Elman 神经网络模型，主要依据有以下两点：一是预测误差尽量小，本研究隐含层神经元节点数为 15 时，模型的平均均方误差在所有模型中最小；二是模型结构尽量简单，本研究预测模型的隐含层节点数仅为 15，其模型的复杂性相对较低，不容易出现过度拟合的现象。

此外，在模型预测因子的选取时发现，流感发病强度与气压、水汽压、气温、SO₂ 浓度、NO₂ 浓度、CO 浓度、PM_{2.5} 浓度、PM₁₀ 浓度、AQI、ILI 病例流感病毒核酸阳性率等因素存在明显相关性，与其他同类研究结论^[17-18]类似，但水汽压和 CO 在以往研究中较少报道阳性结果，其作为预测指标的可靠性尚需更多数据、更长时间的验证。一般认为气象因素、空气质量因素通过影响病毒的生存、人群活动、个体行为等进而对流感流行造成影响^[19]。气象因素、空气质量对流感流行的影响作用，也提示流感预测应更多考虑纳入其他有意义的解释变量，而不应仅局限于 ILI 资料^[3-4,20]。

综上所述，利用气象、空气污染指标和流感病原性监测资料建立的 Elman 神经网络模型，在流感短期预测中具有较好的应用前景。但在应用 Elman 神经网络时，须注意该模型为短期预测模型，其模型参数估计结果依赖所采用的训练数据，连续预测时若出现预测结果误差开始明显变大，应更新数据重新进行训练，建模和预测。

参考文献

[1] PAUL S, MGBERE O, ARAFAT R, et al. Modeling and forecasting influenza-like illness (ILI) in Houston, Texas using three surveillance data capture mechanisms [J]. Online journal of public health informatics, 2017, 9 (2): e187.
 [2] 谢亮, 富小飞, 元云鹏, 等. 嘉兴市流行性感冒监测分析 [J]. 预防医学, 2017, 29 (8): 806-808.

- [3] 陈健, 张磊, 陆帅, 等. 采用广义可加模型分析预测上海市流感样病例发病情况 [J]. 中华流行病学杂志, 2013, 34 (4): 404-408.
- [4] 杨召, 叶中辉, 赵磊, 等. ARIMA-BPNN 组合预测模型在流感发病率预测中的应用 [J]. 中国卫生统计, 2014, 31 (1): 16-18.
- [5] 王斌, 冯慧芬, 黄平, 等. 人工神经网络模型在预测重症手足口病中的应用研究 [J]. 现代预防医学, 2018, 45 (11): 1921-1924, 1947.
- [6] NIU Z, JIA Y, ZHANG L, et al. Prediction for short-term traffic flow based on Elman neural network optimized by CPSO [J]. Metallurgical & Mining Industry, 2015 (9): 991-995.
- [7] CHU C, CHANG C, CHANG T, et al. Elman neural network identify elders fall signal base on second-order train method [C] //2017 6th International Symposium on Next Generation Electronics (ISNE), 2017: 1-4.
- [8] 黄薇, 刘建平, 张锦周, 等. 细菌性食源性疾病发病预测中 Elman 神经网络模型的应用 [J]. 公共卫生与预防医学, 2013, 24 (6): 10-13.
- [9] 孙锦峰, 耿云亮, 郭奕瑞, 等. Elman 神经网络与 ARIMA 模型对流感发病率预测效果的比较 [J]. 郑州大学学报 (医学版), 2013, 48 (5): 584-587.
- [10] 中华人民共和国卫生部. 关于印发《全国流感监测方案 (2010年版)》的通知 [Z]. (2010-09-10) [2018-08-24]. <http://www.nhfp.gov.cn/jkj/s3577/201009/3fa356d0f4834d408fde6c12891a6482.shtml>.
- [11] 何树红, 杨博, 戴明爽, 等. 基于动态递归神经网络的石油价格预测 [J]. 云南民族大学学报 (自然科学版), 2013, 22 (1): 33-35.
- [12] 黎漓. 湖南省流行性腮腺炎的流行特征及 Elman 神经网络预测 [D]. 衡阳: 南华大学, 2015.
- [13] 王芳, 涂春丽, 勾永尧. 基于 Elman 神经网络的气温预测研究 [J]. 安徽农业科学, 2011, 39 (33): 20859-20860.
- [14] KOELLE K, RASMUSSEN D A. Influenza: prediction is worth a shot [J]. Nature, 2014, 507 (7490): 47-48.
- [15] 邓斐, 资海荣, 余慧燕, 等. 江苏省 2010—2014 年流感流行特征分析 [J]. 中华疾病控制杂志, 2016, 20 (9): 901-904.
- [16] 李明, 冯录召, 曹玉, 等. 中国 2005—2013 年流感暴发疫情的流行病学特征分析 [J]. 中华流行病学杂志, 2015, 36 (7): 705-708.
- [17] 王伟. 2009—2014 年山东省流感样病例流行特征及其与气象因素关系研究 [D]. 济南: 山东大学, 2016.
- [18] 杨旭辉, 王婧, 孙昼, 等. 气象因子与流感样病例的相关性研究 [J]. 浙江预防医学, 2014, 26 (7): 653-658.
- [19] SOEBIYANTO R P, CLARA W A, JORGE J, et al. Associations between seasonal influenza and meteorological parameters in Costa Rica, Honduras and Nicaragua [J]. Geospatial Health, 2015, 10 (2): 372.
- [20] 许国章, 王春丽, 李永东, 等. 2008—2015 年宁波市流感样病例预测模型分析 [J]. 国际流行病学传染病学杂志, 2016, 43 (1): 30-34.

收稿日期: 2018-08-24 修回日期: 2018-11-25 本文编辑: 徐文璐

(上接第 112 页)

- [16] 于洋, 郑浩, 费娟, 等. 2012 年江苏省农村环境卫生调查 [J]. 环境与健康杂志, 2013, 30 (8): 709-711.
- [17] 王欣梅, 黄涛, 陈艳华. 2012—2014 年湖南省农村环境卫生健康危害因素调查 [J]. 环境与健康杂志, 2016, 33 (10): 878-902.
- [18] 马李, 王志强, 冯晓亮, 等. 2013 年安徽省农村环境卫生健康危害影响因素调查 [J]. 热带病与寄生虫学, 2014, 12 (3): 137-145.
- [19] 李洪兴, 陶勇, 刘开泰. 我国改水与环境卫生干预控制腹泻病发病效果的 Meta 分析 [J]. 环境与健康杂志, 2014, 31 (5): 438-441.
- [20] 印悦, 金立坚, 秦岭, 等. 四川省农村夏季腹泻发病的环境影响因素 [J]. 环境与职业医学, 2013, 30 (6): 401-404.

收稿日期: 2018-07-06 修回日期: 2018-09-21 本文编辑: 姜安琪