

BP 神经网络模型用于气象因素对脑出血死亡影响的初步研究

高嵩璐 兰莉 乔冬菊 赵娜 杨佳琦 邵冰 焦喆 李航 王滨有

【摘要】 目的 探讨 BP 神经网络预测模型在分析气象因素与脑出血死亡率关系中的应用。方法 根据 BP 神经网络的特性,利用 MATLAB 7.0 软件的神经网络工具箱对 2007—2009 年哈尔滨市气象数据建立脑出血死亡率的 BP 神经网络预报模型,并与传统的多元线性回归模型进行比较。结果 利用多元线性回归结果显示脑出血死亡率与最高气温、最小相对湿度负相关,与平均相对湿度、日照时数呈正相关。脑出血死亡率的非线性相关系数(RNL)为 0.7854,平均绝对误差百分比(MAPE)为 0.21,均方误差(MSE)为 0.22,平均绝对误差(MAE)为 0.19,预测准确度(P)为 81.31%,平均误差率为 0.19。BP 神经网络模型的拟合结果显示,脑出血死亡率的 RNL 为 0.7967,MAPE 为 0.19,MSE 为 0.21,MAE 为 0.18,P 为 82.53%,平均误差率为 0.17。结论 应用 BP 神经网络预测模型对 2010 年哈尔滨市脑出血死亡率进行预报,通过与多元线性回归模型预报结果进行比较,表明该模型具有更高的预报准确度。

【关键词】 BP 神经网络; 脑出血; 气象; 预测

A preliminary study on the effects of meteorological factors on intracerebral hemorrhage death using the BP neural network model GAO Han-lu^{1,2}, LAN Li², QIAO Dong-ju¹, ZHAO Na², YANG Jia-qi¹, SHAO Bing¹, JIAO Dong-ju¹, LI Hang¹, WANG Bin-you¹. 1 Department of Epidemiology, School of Public Health, Harbin Medical University, Harbin 150081, China; 2 Division of Chronic and Non-communicable Diseases, Harbin Center for Disease Control and Prevention
Corresponding author: WANG Bin-you, Email: wangbyhd@126.com
This work was supported by a grant from the Research Projects of Department Health in Heilongjiang Province (No. 2009-544).

【Abstract】 Objective Using the Back Propagation (BP) Neural Network Model to discover the relationship between meteorological factors and mortality of intracerebral hemorrhage, to provide evidence for developing an intracerebral hemorrhage prevention and control program, in Harbin. Methods Based on the characteristics of BP neural network, a neural network Toolbox of MATLAB 7.0 software was used to build Meteorological data of 2007-2009 with intracerebral hemorrhage mortality to predict the effect of BP neural network model, and to compare with the traditional multivariate linear regression model. Results Datas from the multivariate linear regression indicated that the cerebral hemorrhage death mortality had a negative correlation with maximum temperature and minimum humidity while having a positive correlation with the average relative humidity and the hours of sunshine. The linear correlation coefficient of intracerebral hemorrhage mortality was 0.7854, with mean absolute percentage (MAPE) as 0.21, mean square error (MSE) as 0.22, mean absolute error (MAE) as 0.19. The accuracy of forecasting was 81.31% with an average error rate as 0.19. The Fitting results of BP neural network model showed that non-linear correlation coefficient of intracerebral hemorrhage mortality was 0.7967, with MAPE as 0.19, MSE as 0.21, MAE as 0.18. The forecasting accuracy was 82.53% with the average error rate as 0.17. Conclusion The BP neural network model showed a higher forecasting accuracy when compared to the multiple linear regression model on intracerebral hemorrhage mortality, using the data of 2010's.

【Key words】 BP neural network; Intracerebral hemorrhage; Meteorology; Forecast

脑卒中的发病与环境、饮食生活习惯和气候等

因素有关。我国脑卒中发病的地区分布总体呈现北高南低、西高东低的特征^[1]。目前国内外学者对气象因素与脑血管病关系的研究结论不尽相同^[2-4]。人工神经网络已应用于临床医学各领域^[5-7],但应用 BP 神经网络对出血型脑卒中(脑出血)的研究尚未

DOI: 10.3760/cma.j.issn.0254-6450.2012.09.014

基金项目:黑龙江省卫生厅科研项目(2009-544)

作者单位:150081 哈尔滨医科大学流行病学教研室(高嵩璐,杨佳琦,邵冰,焦喆,李航,王滨有);哈尔滨市疾病预防控制中心慢病所(高嵩璐,兰莉,乔冬菊,赵娜)

通信作者:王滨有, Email: wangbyhd@126.com

见报道。因此,本研究收集 2007—2010 年哈尔滨市脑出血死亡资料及同期气象要素资料,分析其发病特征及其与气象要素的关系,并建立 BP 神经网络预报模型。

资料与方法

1. 资料:脑出血死亡数据来源于 2007—2010 年全国死因登记报告信息系统中哈尔滨市 6 个区(道里、南岗、道外、平房、松北、香坊)的统计,汇总成以周为单位死亡人数,计算周死亡率。同期人口资料来源于哈尔滨市所辖市区的户籍管理部门,其中 2007 年人口数为 3 459 928 人,2008 年人口数为 3 495 232 人,2009 年人口数为 3 525 691 人,2010 年人口数为 3 731 560 人。气象资料来源于哈尔滨市气象局,选取该市 2007—2010 年逐日最高气压(Pa)、最低气压(Pa)、最高气温(°C)、最低气温(°C)、平均相对湿度(%),最小相对湿度(%),降水量(mm)、日照时数(h)进行汇总、统计。对汇总后的周气象资料数据与脑出血死亡率进行多元线性回归分析。

2. BP 神经网络模型的建立:

(1) 变量筛选原则:以多元线性回归分析筛选出的 2007—2009 年周气象数据(最高气温、平均相对湿度、最小相对湿度、日照时间)作为 BP 神经网络训练集的输入,同时期脑出血周死亡率作为输出,建立 BP 神经网络模型;将 2010 年周气象数据作为 BP 神经网络预测集的输入,2010 年脑出血周死亡率作为输出。用 Excel 软件将输入、输出矩阵导入 MATLAB 7.0 软件。

(2) 网络拓扑结构:本文采用含有一层隐含层的三层 BP 神经网络。输入向量个数由输入向量维数决定,本文输入神经元个数为 4。采用试凑法确定隐层神经元个数,先设置较少的隐结点加以训练网络,然后逐渐增加隐节点数,从中确定网络误差最小即是对应的隐节点数,由此隐层神经元的个数选定为 6。输出层神经元个数的选择在本文中为脑出血死亡率,因此神经元个数为 1。本研究采用 MATLAB 软件神经网络工具箱中的前处理函数 [pnmimp, maxp, tn, mint, maxt] = premmmx(input_train, output_train) 对训练样本(input_train, output_train)以及目标样本(input_test, output_test)进行归一化处理。采用神经网络工具箱函数 postmnmx 对归一化后数据的运算结果进行反归一化。

(3) 网络传递函数和参数的选择:本研究中隐

层神经元的传递函数为 tansig, 输出层神经元的传递函数为 purelin, BP 神经网络的学习参数选用 Levenberg-Marquardt 方法(trainlm)。为了防止过拟合现象,本文对数据不断调整网络参数,通过训练发现将网络最大训练次数设为 100 次,训练步长为 50,学习速率为 0.01,动量系数为 0.9,训练目标误差为 0.1,可达到最佳训练效果。

(4) 网络效果的验证:为验证网络的泛化能力,使用 2010 年周气象资料作为验证样本的输入,比较预测结果与真实值之间的差异。

结果

1. 脑出血死亡情况:对 2007—2010 年哈尔滨市脑出血病例资料进行统计,其中冬春季(1、4、10 和 12 月)死亡人数较高,夏季(6—8 月)相对较少。发病年龄特征和死亡季节变化见表 1 和图 1。

表 1 2007—2010 年哈尔滨市脑出血死亡病例年龄特征

年份	男性		女性		合计	
	人数	年龄(岁, $\bar{x} \pm s$)	人数	年龄(岁, $\bar{x} \pm s$)	人数	年龄(岁, $\bar{x} \pm s$)
2007	1241	62.60 ± 14.09	686	67.66 ± 12.97	1927	64.40 ± 13.91
2008	1259	62.35 ± 13.99	694	68.71 ± 13.09	1953	64.61 ± 14.01
2009	1264	62.04 ± 14.03	676	68.16 ± 12.83	1940	64.17 ± 13.93
2010	1289	61.52 ± 14.09	691	68.19 ± 12.84	1980	63.85 ± 13.03
合计	5053	62.12 ± 14.05	2747	68.19 ± 12.93	7800	64.26 ± 13.97

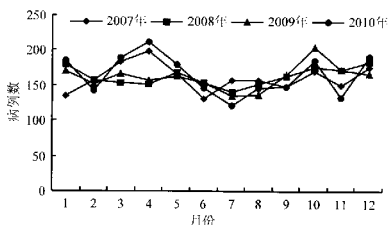
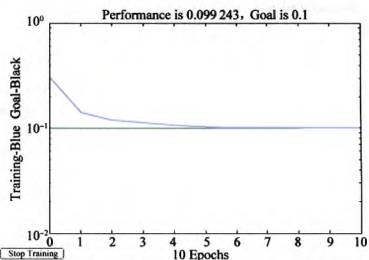


图 1 2007—2010 年哈尔滨市脑出血死亡病例时间分布

2. 多元线性回归分析:将 2007—2009 年哈尔滨市周气象资料[最高气温 X_1 、最低气温 X_2 、最高气温 X_3 、最低气温 X_4 、平均相对湿度 X_5 、最小相对湿度 X_6 、降水量 X_7 、日照时数 X_8]作为自变量,2007—2009 年脑出血周死亡率作为因变量,进行多元线性回归分析,运用逐步后退法,sls=0.15,最终将最高气温、平均相对湿度、最小相对湿度、日照时数纳入回归自变量型。得出多元线性回归模型为: $Y = 1.006 21 - 0.006 15X_1 + 0.005 04X_2 - 0.006 78X_3 + 0.002 16X_4$ ($P_{X_1} < 0.0001, P_{X_2} = 0.0808, P_{X_3} = 0.0115, P_{X_4} = 0.1351$)。

3. BP神经网络预测效果与多元线性回归预测结果比较: 预测网络模型经过10次训练后达到预先设定的最小误差值0.1, 训练情况见图2。



注: 横坐标为步长, 纵坐标为训练误差(0.099243), 达到误差0.1允许范围内

图2 BP神经网络训练曲线

将预测检验数据导入训练好的神经网络中, BP神经网络预测误差为2010年哈尔滨市脑出血实际死亡率与预测死亡率之差, 由图3可知预测误差在10%~30%之间, 预测效果较好。

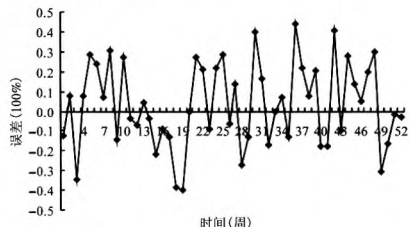


图3 BP神经网络预测误差

将2010年哈尔滨市气象资料代入多元线性回归方程, 得出2010年哈尔滨市脑出血死亡率预测结果, 两种预测结果与实际死亡率比较可知, 绝大部分BP神经网络预测误差较多元回归预测误差小(表2)。利用SAS 9.1软件, 采用*t*检验统计方法($\alpha=0.05$), $F=3.81, P<0.0001$ 。结果表明两种方法预测脑出血死亡率有差别, BP神经网络预测误差较小。

由图4可见, BP神经网络模型在预测极值上比传统多元线性回归更有优势, 预测结果更接近实际死亡率。不同指标对两种模型预测结果质量评价见表3。

讨论

BP神经网络性能受学习样本大小的影响。学

表2 2010年哈尔滨市脑出血死亡率与实际死亡率两种方法预测比较

时间(周)	实际死亡率(10万)	死亡率预测值(10万)		预测误差绝对值(10万)	
		多元线性回归	BP神经网络	BP神经网络	多元线性回归
1	1.2595	1.0356	1.1333	0.1262	0.2239
2	2.10183	1.0666	1.0932	0.0749	0.0482
3	1.4471	1.0723	1.1024	0.3447	0.3749
4	1.0183	1.1451	1.0939	0.0756	0.1268
5	0.7772	1.2015	1.0654	0.2882	0.4244
6	0.8843	1.1324	1.1241	0.2398	0.2481
7	0.9915	1.1739	1.0652	0.0737	0.1824
8	0.8040	1.0939	1.1097	0.3057	0.2899
9	1.2059	1.1620	1.0642	0.1417	0.0439
10	0.9647	1.1244	1.2348	0.2701	0.1596
11	1.2327	1.1291	1.1985	0.0342	0.1036
12	1.2595	1.1032	1.1887	0.0708	0.1563
13	1.2059	1.1167	1.2523	0.0464	0.0892
14	1.1791	1.0922	1.1407	0.0384	0.0870
15	1.3667	1.1547	1.1472	0.2195	0.2120
16	1.4739	1.1717	1.3806	0.0933	0.3023
17	1.2327	1.0424	1.1046	0.1281	0.1904
18	1.4739	1.0276	1.0907	0.3832	0.4463
19	1.3131	1.0071	0.9109	0.4022	0.3061
20	1.0451	1.0183	1.0385	0.0066	0.0269
21	0.8576	1.0726	1.1268	0.2692	0.2151
22	0.8308	1.0403	1.0425	0.2117	0.2096
23	0.9915	1.0280	0.8993	0.0922	0.0365
24	0.7504	0.9702	0.9707	0.2203	0.2198
25	0.8576	1.0669	1.1449	0.2873	0.2093
26	1.0183	0.9503	0.9560	0.0623	0.0680
27	0.5896	0.8970	0.7299	0.1403	0.3074
28	1.0987	0.9824	0.8239	0.2748	0.1164
29	0.8308	0.8316	0.7020	0.1288	0.0009
30	0.4556	0.9675	0.8540	0.3984	0.5119
31	0.6968	0.9276	0.8637	0.1669	0.2309
32	0.9915	0.9682	0.8218	0.1697	0.0233
33	0.8576	0.9873	0.8543	0.0033	0.1298
34	0.8843	1.0138	0.9521	0.0678	0.1295
35	0.9379	0.9641	0.8053	0.1326	0.0262
36	0.6700	1.0825	1.1092	0.4392	0.4126
37	0.8040	1.0649	1.0214	0.2174	0.2610
38	1.1523	1.1094	1.2288	0.0765	0.0429
39	0.9647	1.0664	1.1669	0.2022	0.1017
40	1.3667	1.0628	1.1878	0.1789	0.3039
41	1.2327	1.0228	1.0556	0.1771	0.2099
42	0.9111	1.1415	1.3203	0.4092	0.2304
43	1.1523	1.0610	1.0563	0.0960	0.0913
44	0.8843	1.1129	1.1616	0.2773	0.2286
45	0.8843	1.0497	1.0186	0.1343	0.1654
46	0.7772	1.0143	0.8283	0.0511	0.2372
47	0.8308	1.0210	1.0314	0.2006	0.1903
48	0.8040	1.1182	1.1043	0.3003	0.3143
49	1.3935	1.0748	1.0875	0.3060	0.3188
50	1.3131	1.1209	1.1482	0.1649	0.1922
51	0.9915	1.0472	0.9745	0.0170	0.0557
52	1.1255	1.0400	1.0952	0.0303	0.0855

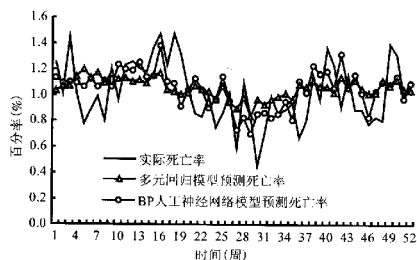


图4 2010年哈尔滨市脑出血死亡率BP神经网络模型预测值、多元回归模型预测值与实际值比较

表3 不同指标对两种模型预测结果质量评价

指标	多元线性回归预测(%)	BP神经网络预测(%)
MAPE	20.92	19.39
MSE	22.47	21.29
MAE	19.07	17.82
P	81.31	82.53
RNL	78.54	79.67
平均误差率	18.69	17.47

注:平均绝对误差百分比(MAPE) = $\frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{|y_i|} \times 100\%$; 均

方误差(MSE) = $\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \times 100\%$; 平均绝对误差(MAE) =

$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \times 100\%$; 预测准确度(P) = $\left[1 - \frac{MAE}{\frac{1}{n} \sum_{i=1}^n y_i} \right] \times$

100% ; 非线性相关系数(RNL) = $1 - \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2}} \times 100\%$; 平

均误差率 = 平均误差绝对值/实际均值; y_i : 实际死亡率, \hat{y}_i : 预测死亡率

习样本增加,则网络更容易收敛到全局最优,且其泛化能力也逐渐提高。学习样本增大到一定程度,则隐层节点的数目对网络影响不大,学习精度对网络的影响也不大,而且网络很稳定,预测效果也很好^[7]。通过多元线性回归筛选出的气象资料信息作为BP神经网络模型的输入信息,设计合理可行,能满足ANN模型的训练和数据的要求,在本文中基于BP神经网络构建的模型其自学习能力和非线性处理能力的优势,使其比传统多元回归模型具有更高的准确度,且误差较传统预报模型更小,所以BP神经网络模型可作为一种简便易行、非传统统计的方

法应用到医学气象预报领域,从而减少脑出血的死亡,延长患者寿命,提高患者生活质量的目的,为脑出血死亡率提供了新的预报方法。但是由于气象因素只是脑出血死亡的一个诱发因素,脑出血死亡与发病时间不相等等因素可能导致预测与实际结果有偏差。为了进一步提高对脑出血死亡的预测准确度,可选择与脑出血相关的关键气象指标和个体因素(如年龄、职业、疾病史等)作为网络输入,或增加其他环境因素及其因素间的交互作用作为神经网络的输入,可能会有助于改进神经网络的训练和泛化能力,提高脑出血死亡率的预报。

参 考 文 献

- [1] Wang WZ. Neurology. Beijing: People's Medical Publishing House, 2008: 122. (in Chinese)
王维治. 神经病学. 北京: 人民卫生出版社, 2008: 122.
- [2] Gonçalves FL, Braun S, Dias PL, et al. Influences of the weather and air pollutants on cardiovascular disease in the metropolitan area of Sao Paulo. Environ Res, 2007, 104: 275-281.
- [3] Barnett AG. Temperature and cardiovascular deaths in the US elderly: changes over time. Epidemiology, 2007, 18: 369-372.
- [4] Zhao KL, Liu CP, Meng LJ, et al. Effect of meteorological factors on the incidence of cerebral hemorrhage in Shijiazhuang area residents. Modern Journal of Integrated Traditional Chinese and Western Medicine, 2009, 28: 3446-3447. (in Chinese)
赵科林, 刘彩平, 孟玲洁, 等. 气象因素对石家庄地区居民脑出血发病的影响. 现代中西医结合杂志, 2009, 28: 3446-3447.
- [5] Qin P, Zhang LZ, Zhao XW, et al. Application of back-propagation neural network in analysis of influencing factors for metabolic syndrome. Pract Prev Med, 2011, 18: 1819-1822. (in Chinese)
秦平, 张豫豫, 赵晓雯, 等. BP神经网络在代谢综合征影响因素分析中的应用. 实用预防医学, 2011, 18: 1819-1822.
- [6] Gao XJ. Application of BP neural network in prediction of the liver cirrhosis's cure. Math Theory Appl, 2011, 31: 20-23. (in Chinese)
高秀娟. BP神经网络在肝硬化治疗预测中的应用. 数学理论与应用, 2011, 31: 20-23.
- [7] Zhang J, Huang ZJ. Principle of BP neural network model and its application in the field of psychology. Mod Prev Med, 2006, 33: 1854-1857. (in Chinese)
张军, 黄子杰. BP神经网络模型的原理及在心理学领域的应用. 现代预防医学, 2006, 33: 1854-1857.

(收稿日期: 2012-03-19)

(本文编辑: 张林东)