

· 基础理论与方法 ·

智能神经网络模型用于气象因素对疟疾发病影响的初步研究

高春玉 熊鸿燕 易东 柴光军 杨小为 刘莉

【摘要】 目的 建立气象因素与疟疾的智能神经网络预测模型。方法 利用 Matlab 6.1 软件中的神经网络工具箱 根据预实验结果 利用云南省红河地区 1994~1999 年月平均气压、月平均气温、月最高气温、月最低气温、月降水量、月降水日数、月平均相对湿度、月蒸发量、月日照时数等气象数据与疟疾发病率等级数据建立反向传播网络(BP 网络)预测模型 并对模型进行验证。结果 神经网络经 100 次学习和训练 训练误差从 3.236 08 下降至 0.035 862 通过建立的智能神经网络模型对未来疟疾发病率进行预测 其预测符合率为 84.85%。结论 智能神经网络在气象因素与疟疾之间建模是可行的 其预测符合率达 80% 以上。智能神经网络具有综合能力强 对数据的要求不高 适时学习等突出优点 且操作简便 节省时间 易于掌握和应用。研究中数据的应用、纳入、排出等问题有待于进一步研究。智能神经网络模型可以作为疟疾预测的一种新方法。

【关键词】 疟疾; 气象因素; 神经网络模型; 流行病学

Study on meteorological factors-based neural network model of malaria GAO Chun-yu*, XIONG Hong-yan, YI Dong, CHAI Guang-jun, YANG Xiao-wei, LIU Li. *Airforce Center for Diseases Control, Sanitation and Anti-epidemic Team, Logistics Department, Airforce, Beijing 100076, China

【Abstract】 Objective In order to provide reliable data for strategies development on prevention, a meteorological factors-based predicating model for malaria forecast was studied. **Methods** Data on malaria occurrence and climate changes from 1994 to 1999 in counties in Yunnan province was collected and analyzed with software packages of FoxPro 6.0 and Excel 5.0. The forecasting model for malaria occurrence was established, using the Neural Network Toolbox of Matlab 6.1 software package. In the studies of forecasting model, data of malaria and meteorological factors from 1994 to 1999 in Honghe state in Yunnan province was chosen. The meteorological factors included average monthly pressure, air temperature, relative humidity, monthly maximum air temperature, minimum air temperature, rainfall, rainday, evaporation and sunshine hours in the study. The established forecasting model was also tested and verified. **Results** The BP network model was established according to data of diseases and meteorological factors from Honghe state in Yunnan province. After training the neural network for 100 times, the error of performance decreased from 3.236 08 to 0.035 862. Verified by fact data of malaria, the efficiency of malaria forecasting was 84.85%. **Conclusion** Neural network model was effective for forecasting malaria. It showed advantages as: strong ability for analysis, lower claim for data, convenient and easy to apply etc. Neural network model might be used as a new method for malaria forecasting.

【Key words】 Malaria; Meteorological factors; Neural network model; Epidemiology

疟疾是全世界最重要的虫媒传染病之一。近年来研究发现 疟疾发病率上升、流行季节延长、流行疟区扩大的趋势与全球气候异常变暖存在密切关系^[1-5]。不少学者试图在气象因素与疟疾之间建立数学模型^[6-13] 预测在变化的气候中疟疾的流行趋势 但这些模型存在着或是建模程序复杂 或是对数

据条件限制多、或是预测准确度不够理想等问题。因此探索和建立新的预测模型技术是非常必要的。近 10 年来,智能神经网络(神经网络)技术作为人工智能技术的新手段得到较大发展,目前神经网络已被应用到科学计算、经济预测、安全性评价、自动控制等众多领域^[14-18],但用于疾病预测方面的神经网络模型尚未见报道。为探讨气象因素对疟疾流行的影响规律,建立气象因素与疟疾的神经网络预测模型,在变化的气候中预测疟疾的发病状况,为制定疟

作者单位:100076 北京 空军后勤部卫生防疫队疾病监控室(高春玉、柴光军、杨小为、刘莉);第三军医大学流行病学教研室(熊鸿燕);第三军医大学统计学教研室(易东)

疾预防控制措施提供科学依据。我们收集了1994~1999年云南省各县市疟疾发病资料、云南省国家气象观测站的各种气象资料,利用 Matlab 6.1 软件中的神经网络工具箱(neural network toolbox, NNT)^[19]在气象因素与疟疾之间初步建立了神经网络模型并用模型进行预测。

基本原理

1. 研究方法:

(1) 选用模型:反向传播网络(backpropagation network, BP 网络)模型由若干层神经元组成,可分为输入层、隐层和输出层。网络中的信息传递是单向的,同一层中的神经元之间无联系,而层与层之间多采用全互联方式,其连接程度用权值表示并可通过学习来调节其值^[20],通过训练,它能够很好地模拟任何有有限断点的函数。

(2) 模型建立流程:数据输入及网络初始化→网络训练→网络的建立→预测模型建立。

(3) 模型成立条件:根据预实验和实践经验,暂定以网络模拟输出值与实际值符合率达到 80% 以上为模型成立条件。

(4) 疟疾发病率分等级即神经元个数:此模型是按发病率的高低进行预测,结合专业知识及当地往年疟疾发病率情况及预实验结果将发病率定为 4 个等级(4 个神经元):月发病率 < 1/10 万定为 1 级,1~<5/10 万定为 2 级,5~<10/10 万定为 3 级,≥ 10/10 万定为 4 级。

2. 模型建立:

(1) 数据输入及网络初始化:用 Matlab 语言编程。

① 输入矩阵 P 表示,其中 A:月平均气压;B:月平均气温;C:月最高气温;D:月最低气温;E:月降水量;F:月降水日数;G:月平均相对湿度;H:月蒸发量;I:月日照时数;n 输入变量个数。

P=[A1 1	A1 2	A1 3	A1 n ;
B2 1	B2 2	B2 3	B2 n ;
C3 1	C3 2	C3 3	C3 n ;
D4 1	D4 2	D4 3	D4 n ;
E5 1	E5 2	E5 3	E5 n ;
F6 1	F6 2	F6 3	F6 n ;
G7 1	G7 2	G7 3	G7 n ;
H8 1	H8 2	H8 3	H8 n ;
I9 1	I9 2	I9 3	I9 n ;

② 输入目标向量 T 值,即 T=[T1 T2 T3

..... Tn],式中 T 为目标向量即疟疾发病率等级,为与输入元素向量一一对应的疟疾发病率等级。

③ 网络初始化:net = newff([minA maxA ; minB maxB ;..... ;minR maxR],[s 1],{ 'tansig' 'purelin' });式中 net 代表变量,newff 代表网络结构,是一个 R×S 的矩阵,R 表示输入元素的个数,minA maxA minB maxB ;..... minR maxR 分别表示 R 个输入元素的最小值和最大值范围。S 为隐层神经元个数;1 表示输出层为 1 个神经元。tansig 为隐层传递函数,purelin 为输出层传递函数。

④ 设定网络迭代次数,即 net.trainParam.epochs = 100 ;net.trainParam.epochs 参数为网络训练步数,即迭代次数,反复调整至训练误差逐步下降至最小时(目标是 0),其输出值与实际值符合率越高。

(2) 网络训练:net = train(net ,P ,T);通过建立的可训练 BP 网络、输入元素向量及目标向量,网络自适应训练和学习。

(3) 结果模拟:a = sim(net ,P);式中 a 为模拟输出值;用模拟函数 sim 接收网络输入 P,网络对象 net,返回网络输出 a。

(4) 预测模型建立:计算网络模拟输出值 a 与实际值 T 符合率,根据模型成立条件,若符合率 > 80% 则预测模型成立。

3. 模型预测验证:

(1) 用上述所建神经网络模型对未来疟疾发病率进行预测,程序如下:

① 输入新变量:

P1=[A1 1	A1 2	A1 3 ;
B2 1	B2 2	B2 3 ;
C3 1	C3 2	C3 3 ;
D4 1	D4 2	D4 3 ;
E5 1	E5 2	E5 3 ;
F6 1	F6 2	F6 3 ;
G7 1	G7 2	G7 3 ;
H8 1	H8 2	H8 3 ;
I9 1	I9 2	I9 3];

② 网络预测:forecast = sim(net , P1);式中 forecast 为网络预测函数,通过建立起的网络,利用 sim 函数对未来进行预测。

(2) 计算预测结果:将网络输出预测值 a1 与已知真实值 T 进行比较,以求出预测符合率。

实例分析

1994~1999 年云南省各县市的疟疾月发病数

资料来自于国家疫情报告管理数据库,按照地区、发病月、发病数进行重新整理和建库。云南省国家气象观测站 1994~1999 年月气象资料来自于中国气象局气候资料中心,共收集 31 个观测站资料。按照地区、月平均气压、月平均气温、月最高气温、月最低气温、月降水量、月降水日数、月平均相对湿度、月蒸发量、月日照时数等 9 种气象因素进行重新整理和建库。1994~1999 年云南省各县市人口数由全军疾病监测中心提供。

利用云南省红河哈尼族彝族自治州 3 个国家气象观测站所在县(屏边苗族自治县、蒙自县、泸西县) 1994~1998 年当月平均气象数据与后 1 个月疟疾发病率等级数据建立神经网络模型。利用上述 3 个县 1999 年 1~11 月气象数据预测该地区 1999 年 2~12 月疟疾发病率范围。

1. 气象因素与疟疾的智能神经网络模型建立:

(1) 数据输入及网络初始化:

① 输入矩阵 P 输入上述 3 个县 1994 年 1 月至 1998 年 11 月月平均气压、月平均气温、月最高气温、月最低气温、月降水量、月降水日数、月平均相对湿度、月蒸发量、月日照时数。

```
P = [ 868.9  868.5  865.2  864.2  .....  864.0 ;
      18.1   19.2   23.4   23.3  .....  22.1 ;
      29.9   30.4   33.2   32.8  .....  28.6 ;
        9.2   10.6   14.4   15.3  .....  17.6 ;
      55.5  112.2   23.0   58.4  .....  215.1 ;
        9     16     9     15     .....  25 ;
        66    67    60    72     .....  84 ;
      234.6 218.5 295.5 215.9 ..... 120.8 ;
      240.4 202.9 240.8 134.7 .....  48.7 ;
```

② 输入疟疾发病率等级:输入 3 个县与气象因素相隔 1 个月即 1994 年 2 月至 1998 年 12 月疟疾发病率等级。

```
T = [ 1   1   3   4   .....   3 ];
net = newf([ 0 1 000 ; 0 50 ; 0 50 ; -20 30 ; 0 1 000 ; 0 31 ; 0 100 ; 0 1 000 ; 0 500 ] [ 4 1 ], 'tansig' 'purelin');
net.trainParam.epochs = 100 ;
net = train(net, P, T);
a = sim(net, P);
```

(2) 网络训练:在命令窗口执行运行命令,网络开始学习和训练,由图 1 可见此网络经 100 次学习和训练,误差从 3.236 08 下降至 0.035 862。

(3) 结果模拟:网络训练结束后,在程序运行窗口显示与输入元素相对应的全部模拟输出值,每行最多显示 10 个输出值即 1~4 不等的 10 个数值,网

络模拟输出值 a 与实际值 T 比较,结果符合率为 84.62% (表 1)。

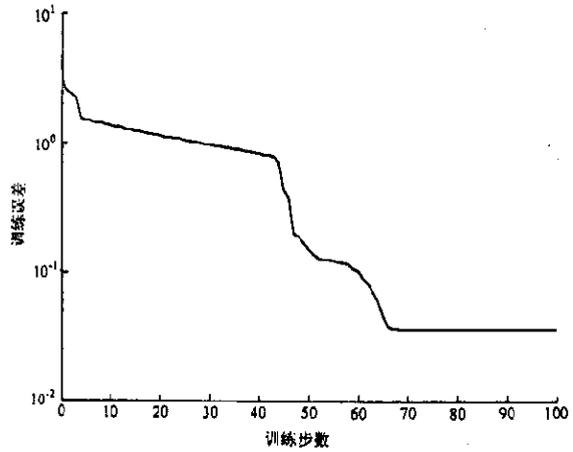


图 1 神经网络训练过程图

表 1 神经网络模拟输出值(a)与实际值(T)比较

发病率 (/10 万)	分级*	T 个数	a 个数	符合数	符合率 (%)
<1	1	85	81	75	88.24
1~	2	34	30	25	73.53
5~	3	21	27	19	90.48
≥10	4	16	18	13	81.25
合计		156	156	132	84.62

* 疟疾发病率分等级

2. 用智能神经网络模型进行预测:

(1) 输入上述 3 个县 1999 年 1 月至 1999 年 11 月的各种气象数据。

```
P1 = [ 866.8  867.3  864.3  863.3  .....  865.6 ;
       21.7   20.7   23.9   23.0  .....  22.0 ;
       32.8   32.4   31.1   30.3  .....  30.3 ;
       13.0   10.5   19.0   18.4  .....  17.8 ;
       56.1   76.5   73.1   218.7 .....  202.8 ;
        9     20     14     19     .....  20 ;
        67    72    14     80     .....  83 ;
      243.3 184.5 193.5 170.8 ..... 132.1 ;
      211.4 131.1 155.2 121.2 .....  94.6 ;
```

```
forecast = sim(net, P1);
```

(2) 预测:程序运行后网络输出预测值,与已知的实际值进行比较,3 个县 1999 年疟疾发病率的预测符合率为 84.85%,其中屏边、蒙自、泸西县预测符合率分别为 81.82%、90.91%、81.82%。

讨 论

神经网络工具箱是 Matlab 中较为重要的一个工具箱,在实际应用中,BP 网络用的最广泛。神经网络具有综合能力强,对数据的要求不高,适时学习

等突出优点,其操作简便,节省时间,网络初学者即使不了解其算法的本质,也可以直接应用功能丰富的函数来实现自己的目的。因此,易于被基层单位预防工作者掌握和应用。研究表明,用此方法建立气象因素与疟疾的神经网络模型是可行的。

在研究中我们体会到,以下几个问题是建立理想的因素与疾病之间的神经网络模型的关键:①资料选取:应尽可能地选取所研究地区系统连续的因素与疾病资料,最好包括有疾病高发年和疾病低发年的数据。在收集影响因素时,要抓住主要影响疾病的因素。②疾病发病率分级:神经网络预测法是按发病率高低来进行预测,在定义发病率等级时,要结合专业知识及当地情况而定,并根据网络学习训练效果而适时调整,以使网络学习训练达到最佳效果。③资料处理:在实践中发现,资料的特征往往很大程度地影响网络学习和训练的稳定性,因此,数据的应用、纳入、排出问题有待于进一步研究。

近年来,建立气象因素与疟疾之间的模型及用模型预测在变化的气候中疟疾流行趋势,引起了许多学者的关注。目前用于疟疾预测的模型主要有:①灰色预测模型:其中 GM(1,1)模型应用较多,此方法较复杂,目前较多用于对多年的发病率来预测发病率,预测效果较理想,但未考虑对发病的影响因素。②灰色关联分析:它是按有关数据序列之间的几何关系及其相似程度来判断其关联程度,通过对母因素和子因素的初值化,根据公式依次计算出第 i 个母因素对第 j 个子因素的 k 时刻的关联系数,计算复杂,不适宜对大量资料的分析^[6]。③通径分析:利用通径图可反映各变量间的直接与间接关系,但通径图中,内生变量与外生变量间的关系都是线性的,实际工作中的非线性偏离被认为可以忽略,如有强的非线性关系应当设法对变量作变换,以便可以用线性作近似,但变量转换很复杂,需反复摸索。此模型常用于检验专业工作者所提出的各种假设、观点或模型的合理性^[8]。④logistic 回归模型、多元回归或线性混合模型:对局部地区疟疾发病率进行分析,预测气候变化对疟疾传播强度的影响有一定的意义^[9],用此模型需要发病与因素间呈线性或分布符合一定规律,但由于疟疾发病与各种气象因素之间的关系复杂,转换起来很复杂而较少使用。

近年来,应用现代技术和数理统计方法研究疟疾的流行有了许多新的尝试,特别是卫星遥感技术结合地理信息系统在疟疾预测及控制方面发挥着重

要作用。神经网络模型如能与卫星遥感技术、地理信息系统结合起来,在疟疾预测及控制方面的作用将会更加明显。相信随着生物学、数学、计算机等众多相关学科的飞速发展,神经网络理论必将在疾病控制领域有广阔的应用前景。

参 考 文 献

- Bouma MJ, Dye C, Kaay VD, et al. Falciparum malaria and climate change in the northwest frontier province of Pakistan. *Am J Trop Med Hyg*, 1996, 55:131-137.
- Freeman T, Bradley M. Temperature is predictive of severe malaria years in Zimbabwe. *Trans R Soc Trop Med Hyg*, 1996, 90:232.
- Rogers DJ, Randolph SE. The global spread of malaria in a future, warmer world. *Science*, 2000, 289:1763-1766.
- Bryan JH, Foley DH, Sutherst RW. Malaria transmission and climate change in Australia. *Med J Aust*, 1996, 164:345-347.
- Loevinsohn ME. Climatic warming and increased malaria incidence in Rwanda. *Lancet*, 1994, 343:714-718.
- 赵天顺, 孙阳, 钟玉仙. 影响传染病流行的主要因素分析和动态模型建立初探. *西南国防医药*, 1998, 8:202-205.
- 方莹, 邓达, 顾政诚, 等. 徐州地区疟疾发病率的区间划分、预测及下降阶段推测模型. *中国寄生虫学与寄生虫病杂志*, 1991, 9:284-286.
- 孙尚拱, 编著. 医学多变量统计与统计软件. 第 1 版. 北京:北京医科大学出版社, 2000. 163-212.
- Kleinschmidt I, Sharp BL, Clarke GP, et al. Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in Kwazulu Natal, South Africa. *Am J Epidemiol*, 2001, 153:1213-1221.
- 巴剑波, 方旭东, 徐雄利. 马尔可夫链在海军疟疾疫情预测中的应用. *解放军预防医学杂志*, 2001, 19:114-116.
- 游本荣, 胡松华. 余弦分析法在疟疾发病季节特征研究中的应用. *数理医药学杂志*, 1995, 8:267-268.
- 王成科, 胡勇. 应用圆形分布研究 5 种疾病发病季节特征. *数理医药学杂志*, 2001, 14:121-123.
- 舒端, 鲁家才, 陈凤莲, 等. 应用圆形分布法分析汉川市疟疾季节性分布特征. *中国寄生虫病防治杂志*, 2001, 14:11.
- Mukher A, Deshpande JM. Application of artificial neural networks in structural design expert systems. *Computer and Structures*, 1995, 54:367-370.
- 闻新, 周露, 王丹力. *Matlab 神经网络应用设计*. 第 1 版. 北京:科学出版社, 2000. 9.
- 弓箭, 刘剑. 基于神经网络的自适应控制应用现状及发展. *信息技术*, 2001, 9:39-41.
- 高英杰, 王海芳, 董国江, 等. *Matlab 神经网络工具箱在系统辨识中的应用*. *液气气动与密封*, 2001, 5:34-36.
- 杨振宏. 基于神经网络理论的安全性预评价方法. *黄金*, 2001, 11:12-15.
- 王学辉, 张明辉, 编著. *Matlab 6.1 最新应用详解*. 第 1 版. 北京:中国水利水电出版社, 2002. 377-402.
- 盛昭昭. 主从递阶决策论. 北京:科学出版社, 1998. 150-155.

(收稿日期:2003-01-03)

(本文编辑:张林东)