

基于广义回归神经网络的赤潮预警

国锋

(青岛酒店管理学院 信息工程技术学院, 山东 青岛 266100)

摘要: 探讨了广义回归神经网络的原理和相关算法, 将广义回归神经网络应用于赤潮预警, 并以米氏凯伦藻为例进行了实验。与目前使用较为广泛的BP神经网络进行比较, 结果表明, 广义回归神经网络的预警效果要优于BP网络, 具有较高的实用价值。

关键词: 赤潮; 预警; 广义回归神经网络

中图分类号: TP391.9

文献标识码: A

Harmful red tide warning based on general regression neural network

GUO Feng

(School of Information Engineering Technology, Qingdao Hismile College, Qingdao 266100, China)

Abstract: This paper discusses the theory and algorithm of general regression neural network, applies it into red tide early warning, and compares general regression neural network with the widely used BP neural network. After research, the results show that general regression neural network in red tide warning is better than the BP neural network, and is worthy for practical use.

Key words: red tide; early warning; GRNN

由于我国沿海地区人口的增长及工农业生产的持续发展, 近海的有机物污染日益严重, 局部海域富营养化问题日益突出^[1], 加之营养盐类大面积超标, 我国沿海有害藻化发生频率急剧增高, 规模和危害程度有愈演愈烈的趋势^[2-3]。当前, 有害赤潮已成为制约我国沿海经济可持续发展的一个重要因素^[4-6], 引起了我国政府的高度重视和社会各界的普遍关注。因此, 搞清我国海域有害赤潮形成机理, 提出预测和防治方法以减轻赤潮灾害, 已成为我国海洋研究工作者的重点课题。

有害赤潮是一种由多因素综合作用引起的生态异常现象, 人们至今还没有对其发生机理有完整、统一的认识, 而且赤潮生物的繁殖与其影响因子之间具有高度的复杂性和非线性, 从而很难采用传统的偏微分方程对其规律进行描述^[2]。在这种情况下, 以数据挖掘技术为主的数据驱动模型就开始显示出其优势。其中, 人工神经网络 ANN (Artificial Neural Network) 的应用最为广泛, 尤其是在寻求和建立信号输入和输出关系方面, 具有许多其他方法所不具备的优势, 并在实际应用中取得了很好的效果。

1 广义回归神经网络 GRNN

人工神经网络是一种简化的人脑数学模型。它由大量具有非线性映射能力的信息处理单元(神经元)广泛连接而成, 是以模拟人脑神经系统结构和功能的复杂网络系统。现有的研究成果表明神经网络已具有人脑的一些基本特征, 如学习、记忆和归纳等功能^[5-6]。

人工神经网络技术有多种模型及相应的学习算法, 例如误差反向传播 BP(Back Propagation)神经网络模型和径向基函数 RBF(Radial Basis Function)神经网络模型等, 并在环境科学与工程的环境质量评价与预测、监测点的优化布置、社会经济环境可持续发展、污染物降解与释放、水处理、生态系统的模拟与预测等方面获得了广泛的应用。本文所采用的是一种新型的人工神经网络——广义回归神经网络 GRNN(General Regression Neural Network)来对赤潮发生有关的物理的、生化的影响因子与赤潮生物的密度之间的关系进行分析和学习, 并利用训练后的 GRNN 网络对赤潮生物的密度进行预测^[7]。

GRNN 网络最早由 Donald F. Specht 在 1991 年提出的, 是一种新型神经网络^[8-9], 它建立在数理统计的基础上, 是一种基于 one-pass 学习算法的一种高度并行的径向基

网络与通信 Network and Communication

网络,其理论基础是非线性(核)回归分析。它能够根据样本数据逼近其中隐含的映射关系,即使样本数据稀少,网络的输出结果也能够收敛于最优回归表面。GRNN网络具有以下优点^[10-12]:

(1)GRNN同样能够以任意精度逼近任意非线性连续函数,且预测效果优于BP网络;

(2)由于GRNN的训练过程是单程训练,不需要迭代,它较BP网络的训练过程快得多,更适合在线数据的实时处理;

(3)GRNN所需的训练样本较BP网络少得多。Specht等人指出,取得同样的预测效果,GRNN所需样本仅为BP网络的1%;

(4)隐层节点数由训练样本自适应确定,网络的可调参数少,网络的学习全部依赖于数据样本,最大程度地减少了主观因素对预测结果的影响。

从网络结构上看,GRNN由一个径向基网络层和一个线性网络层组成,如图1所示。其中, P 表示输入向量, R 表示网络输入的维数,即因子的个数, Q 表示每层网络中神经元的个数,也就是训练样本的个数。 a_i^1 表示第一层输出 a^1 的第 i 个元素, $i/W_{1,j}$ 表示第1层权值矩阵 $W_{1,1}$ (为 $Q \times R$ 矩阵,即全部的样本输入向量)的第 i 行元素。

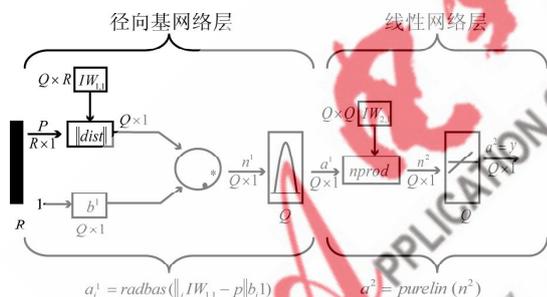


图1 GRNN网络结构图

网络的第一层为径向基隐含层,神经元个数等于训练样本数 Q ,该层的权值函数为欧几里德距离度量函数,即 $\|dist\|_j = \sqrt{\sum_{i=1}^R (x_i - W_{1,i})^2}$ ($j = 1, 2, \dots, Q$),其作用是计算网络输入与第一层的权值 $W_{1,1}$ 之间的距离, b^1 为隐含层阈值。符号“ $\cdot \cdot$ ”表示 $\|dist\|$ 的输出与阈值 b^1 的元素与元素之间的乘积关系,并将结果形成净输入 n^1 ,传递到传递函数。隐含层的传递函数为径向基函数,常用高斯函数 $R_i(x) = \exp(-\|x - c_i\|^2 / 2\sigma_i^2)$ 作为网络的传递函数,式中 x 是 R 维输入向量, c_i 是第 i 个基函数的中心, c 与 x 具有相同的维数, σ_i 决定第 i 个隐含层位置处基函数的形态, σ_i 越大则基函数越平滑,故又被称为光滑因子。如果将网络的输入归一化到区间 $[-1, 1]$ 内,则有 $c_i = 0$,从而得到 $R_i(x) = \exp(-x^2 / 2\sigma_i^2)$,通常将阈值 b^1 设置为满足 $R_i(b_1^1) = 0.5$,此时,当网络的输入与相应权值的距离小于阈值 b_1^1 时,传递

函数具有大于0.5的输出。由 $R_i(b_1^1) = 0.5$ 即 $R_i(b_1^1) = \exp(-(b_1^1)^2 / 2\sigma_i^2) = 0.5$ 得到, $b_1^1 = \sqrt{-\ln 0.5} \cdot \sqrt{2}\sigma_i = 0.8326 \cdot \sqrt{2}\sigma_i$,令 $spread = 1/(\sqrt{2}\sigma_i)$,则有 $b_1^1 = 0.8326/spread$, $spread$ 称为径向基函数的分布系数,当 $spread = 1.0$ 时, $R_i(x) = \exp(-x^2 / 2\sigma_i^2) = \exp(-x^2/spread^2) = \exp(-x^2)$,即为radbas函数。

网络第二层为线性输出层,其权函数为规范化点积权值函数nprod,计算出网络的输出向量 n^2 ,它的每个元素就是向量 a^1 与权值矩阵 $LW_{2,1}$ 每行元素的点积再除以向量 a^1 各元素之和,并将结果 n^2 送入线性传递函数 $a^2 = purelin(n^2)$,计算网络的输出。

GRNN连接权值的学习修正仍然使用BP算法,由于网络隐含层节点中的作用函数(基函数)采用高斯函数,高斯函数是一种局部分布对中心径向对称衰减的非负的非线性函数,对输入信号在局部产生响应,即当输入信号靠近基函数的中央范围时,隐含层节点将产生较大的输出,由此看出这种网络具有局部逼近的能力,这也是该网络之所以学习速度更快的原因。此外,GRNN人为调节的参数少,只有一个 $spread$,网络的学习全部依赖于数据样本。这个特点也决定了网络得以最大限度地避免人为主观假定对预测结果的影响。

2 实验仿真

典型微藻型包括夜光藻、肋条骨藻、塔马亚历山大藻、米氏凯伦藻等。据文献[13-14]所述,董婧等在1997年对海洋岛与王家岛周围海域的调查研究中采集了17个样本,各个样本包含的数据有5个理化因子:水温、溶解氧、盐度、总氮、总磷以及浮游植物密度和夜光4藻的密度。其中浮游植物和米氏凯伦藻(*Noctiluca scientillans*)的监测按照《海洋调查规范》(国家海洋局1975)进行,水文、理化因子的采样及分析按《海洋污染调查暂行规范》(1979)执行。取得的数据如表1所示。

将水温、溶解氧、盐度、总氮、总磷、浮游生物密度6个因子作为网络的6个输入,将米氏凯伦藻的密度作为网络的输出,采用MatLab 6.5的神经网络工具箱分别建立BP网络和GRNN神经网络,并比较二者的预测结果。所采用的BP网络的结构为6-15-1,即隐含层有15个神经元,传递函数采用双曲正切S型函数tansig,训练函数采用Levenberg-Marquardt训练函数,即trainlm,权值和阈值学习函数采用动量批梯度下降函数learngdm,性能函数采用均方误差函数mse,目标迭代次数定为2000。GRNN的 $spread$ 值设置为0.8。一共进行了2次实验,每次随机选取15个样本作为网络的训练数据,并用这15个样本来检查网络的拟合性能,余下2个样本作为预测样本,用以检查网络的性能。得到的结果如表2和图2、图3所示。

表1 米氏凯伦藻密度和各理化因子数据

站位	水温(°C)	溶解氧(mg/L)	盐度	总氮(μmol/L)	总磷(μmol/L)	浮游生物密度(10 ⁴ 个/m ³)	米氏凯伦藻密度(10 ² 个/m ³)
1	23.3	7.22	30.4	0.72	0.10	71.5	530
2	23.2	7.31	30.4	0.73	0.26	160	467
3	23.1	7.50	30.4	0.56	0.37	146	470
4	23.0	7.44	30.6	0.81	0.18	31.1	440
5	23.7	7.36	30.5	2.05	0.36	125	260
6	23.2	7.54	30.5	0.52	0.07	32	180
7	23.4	7.62	30.4	0.88	0.07	31.8	160
8	23.4	7.54	29.5	0.23	0.43	226	60.4
9	22.9	7.56	30.8	0.83	0.21	9.72	23.7
10	24.2	7.83	30.5	0.14	0.40	148	8.28
11	22.5	6.97	30.4	0.52	0.56	187	7.6
12	23.0	7.14	30.3	0.27	0.43	147	7.56
13	23.1	7.12	30.2	0.17	0.40	74.5	5.99
14	23.5	7.38	30.3	0.31	0.47	205	4.04
15	23.0	7.06	30.3	0.40	0.54	109	2.8
16	22.9	7.06	30.3	0.40	0.54	148	1.77
17	24.7	7.56	30.2	0.32	0.32	10.5	91.5

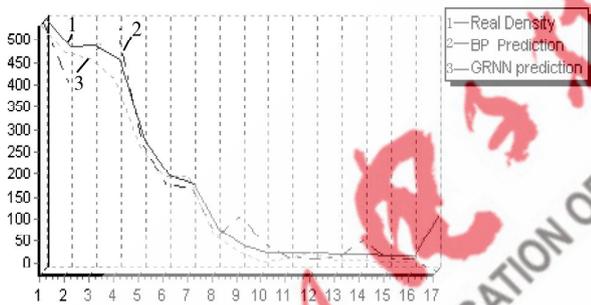


图2 第1次计算值与实测值对比图

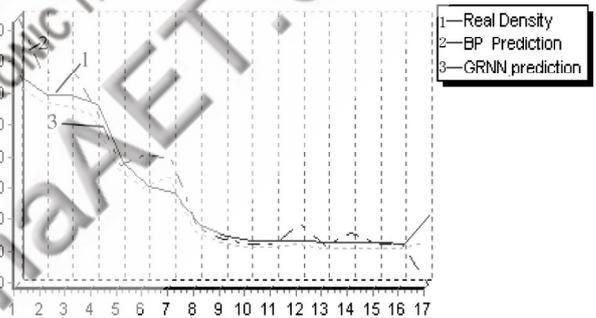


图3 第2次计算与实测值对比图

表2 米氏凯伦藻密度计算值与实测值比较 (单位:10²个/m³)

样本号	实测值	BP 网络预测值	GRNN 预测值	样本号	实测值	BP 网络预测值	GRNN 预测值
1	530	529.992	530.000	1	530	723.593	530.000
2	467	396.267	467.000	2	467	544.026	467.000
3	470	458.010	455.497	3	470	581.451	455.497
4	440	514.294	408.067	4	440	439.997	423.298
5	260	260.004	260.000	5	260	260.000	260.000
6*	180	163.091*	191.871*	6	180	295.547	196.700
7	160	160.031	191.933	7*	160	278.491*	230.472*
8	60.4	50.629	60.400	8	60.4	40.375	60.400
9	23.7	95.469	23.700	9	23.7	23.700	23.700
10*	8.28	34.403*	2.845*	10	8.28	8.109	8.272
11	7.6	4.069	7.600	11	7.6	7.895	7.600
12	7.56	3.043	21.032	12	7.56	72.608	21.029
13	5.99	3.632	5.990	13	5.99	3.780	5.990
14	4.04	39.766	4.040	14	4.04	44.967	4.040
15	2.8	1.988	2.800	15	2.8	7.205	2.800
16	1.77	1.770	2.075	16	1.77	7.248	2.083
17	91.5	-27.764	91.497	17*	91.5	-117.818*	23.700*

注: 带有*的是预测结果

从上面的计算结果可以看出, 两种神经网络都能够较好地对米氏凯伦藻的实际密度进行拟合和预测, 但是 GRNN 网络的精度要明显优于 BP 网络。GRNN 网络的拟合结果中有大量的值和实测值完全相同, 如在第 1 次实验中的第 1、2、5、8、11、13、14、15 号样本和第 2 次实验中的 1、2、5、8、9、11、13、14、15 号样本, 而 BP 网络的拟合结果或多或少总有一些误差。在 2 次预测(外推)的结果中(第 1 次实验的第 6、10 号样本和第 2 次实验的第 7、17 号样本), GRNN 的预测误差要明显小于 BP 网络, 特别是 2 次实验中的第 17 号样本, BP 网络的计算结果出现了负值, 属于明显的失误。

BP 网络预测结果不够准确的原因一方面是由于训练的数据量不足, BP 网络需要用大量的数据来训练才能有比较好的预测效果; 另一方面, BP 网络的参数很多, 许多参数并没有公认的最优值, 很多情况都是靠人们的经验和主观判断, 需要不断地调整、试验才能

网络与通信

Network and Communication

取得比较理想的效果。而 GRNN 网络正好克服了 BP 网络的这些缺点, GRNN 只需要很少的训练样本就可以得到很好的拟合效果和预测效果,同时, GRNN 网络只有 *spread* 一个可调参数,网络的学习全部依赖于数据样本,可以把人为的主观影响降到最低。

另外,在计算的过程中, BP 网络的训练过程是一个迭代过程,需要进行大量的迭代,在本例中, 2 000 次的迭代过程大约需要 40 s 的时间,如果增加样本量或是增加输入因子的个数就需要更长的迭代时间。而 GRNN 网络的训练过程是一个 one-pass 的过程,并不需要迭代,训练几乎是在瞬间完成,特别适合像赤潮预警这种大量实时数据的处理。

可见,人工神经网络是目前在赤潮预警中比较有效的方法,而且相较于 BP 网络, GRNN 网络在赤潮的预警中更具优势。海洋系统是个极其复杂的非线性系统,影响赤潮产生和消亡的因子很多,各因子之间的关系极其复杂,人们至今仍没有统一、完整的认识,在这种情况下,将人工神经网络与传统的分析方法相结合是赤潮预警研究中的一个发展方向。

参考文献

- [1] 国家海洋局.中国海洋灾害公报[EB/OL].<http://www.soa.gov.cn/>, 2005.
- [2] 霍文毅,郝建华,俞志明,等.有害赤潮数值分析研究进展[J].海洋与湖沼,1999,30(5):568-574.
- [3] 潘克厚,姜广信.有害藻华(HAB)的发生、生态学影响和对策[J].中国海洋大学学报,2004,34(5):781-786.
- [4] 吴玉霖,周成旭,张永山,等.烟台四十里湾海域红色裸甲藻赤潮发生过程及其成因[J].海洋与湖沼,2001,32(2):159-167.[5]
- [5] 冯天瑾.神经网络技术[M].青岛:青岛海洋大学出版社,1999.
- [6] 焦李成.神经网络系统理论[M].西安:西安电子科技大学出版社,1996.
- [7] 杨建强,罗先香,丁德文,等.赤潮预测的人工神经网络方法初步研究[J].海洋科学进展,2003,21(3):318-323.
- [8] SPECHT D F. A general regression neural network[J].IEEE Trans Neural Networks,1991(2):568-576.
- [9] SPECHT D F. The general regression neural network rediscovered [J]. IEEE Trans Neural Networks,1993(6):1033-1034.
- [10] 沈掌泉,孔繁胜.基于广义回归网络的动态权重回归神经网络集成方法研究[J].计算机应用研究,2005(12):41-43.
- [11] 周志华,陈世福.神经网络集成[J].计算机学报,2002,25(1):1-8.
- [12] 马慧敏,宋雨,许正伟.广义回归神经网络在软件质量预测中的应用[J].计算机工程与应用,2004(29):217-219.
- [13] 董婧,刘海映,李培军等.海洋岛与王家岛周围海域赤潮生物米氏凯伦藻生态初探[J].海洋环境科学,1999,18(4):48-52.
- [14] 蔡如钰.基于人工神经网络的米氏凯伦藻密度预测模型[J].中国环境监测,2001,17(3):52-55.

(收稿日期:2009-06-24)