

农业科学

ELM 在虫害预测中的应用研究

韦艳玲

(柳州职业技术学院电子信息工程系, 柳州 545006)

摘要 ELM(极限学习机)简单、易用,学习速度快且泛化性好,故将 ELM 引入虫害预测中。为了更好地提高 ELM 的效率和精度,首先采用模糊聚类对所有样本进行预处理,再把处理后的数据作为 ELM 的输入数据进行训练和预测。仿真实验结果表明,经过模糊聚类处理后,ELM 预测精度较高,泛化性好,能够满足虫害预测对准确率和实时性的要求。通过实例,为虫害预测的应用提供一种新方法。

关键词 ELM(extreme learning machine) 模糊聚类 虫害 预测

中图法分类号 S126; **文献标志码** A

传统的虫害预测预报方法之一是线性方法,但在大多数情况下,预报因子与虫害之间常常是非线性关系。人工神经网络由于具有很好的非线性逼近能力,故广泛应用于虫害预测^[1-3]。ELM(extreme learning machine)一般称极限学习机,是一种简单、易用、有效的单隐层前馈神经网络 SLFN 学习算法,由南洋理工大学黄广斌副教授于 2006 年提出。传统的神经网络学习算法(如 BP 算法)需要人为设置大量的网络训练参数,且易产生局部最优解。ELM 只需要设置网络的隐层节点个数,在算法执行过程中不需要调整网络的输入权值以及隐元的阈值,且产生唯一最优解,因此具有学习速度快且泛化性能好的优点^[4]。近年来,有的学者为解决桥梁结构健康监测系统中数据丢失问题,采用 ELM 实现丢失数据的恢复^[5];有的学者为优化文档的相似性计算,研究基于 ELM 的 XML 文档分类方法^[6];有的学者将 ELM 引入到储层渗透率的预测中^[7];有的学者采用 ELM 检测食醋中可溶性无盐固形物含量(SSFSC)的精度和稳定性^[8]等,都取得了令人满意的结果,但极少见到 ELM 应用于虫害预测的公开报道。本文旨在通过实例,为虫害预测的应用提

供一种新方法。

1 ELM 基本原理

ELM 算法是单隐层前馈神经网络(SLFN)的训练算法。设有 N 个样本 (x_i, t_i) , $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in R^n$, $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R^m$, $T = [t_1, t_2, \dots, t_N]$,又设隐含层有 L 个节点,输入节点与第 i 个隐节点的连接权值为 $\omega_i = [\omega_{i1}, \omega_{i2}, \dots, \omega_{in}]^T$,第 i 个隐节点与输出节点的连接权值为 $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]^T$, b_j 是第 i 个隐层节点的阈值。则激活函数为 $g(x)$ 的 ELM 模型的数学表示为^[4]

$$\sum_{i=1}^L \beta_i g_i(x_j) = \sum_{i=1}^L \beta_i g(\omega_i \cdot x_j + b_i) = t_j; \\ j = 1, 2, \dots, N.$$

当激励函数 $g(x)$ 无限可微时,输入连接权值和隐层节点阈值在训练开始时可随机选择,且在训练过程中固定不变。而隐节点与输出节点的连接权值 β 可通过求解以下线性方程组的最小二乘解来获得: $\min_{\beta} \|H\beta - T'\|$ 。

其解为: $\hat{\beta} = H^+ T'$ 。

其中 H^+ 为隐层输出矩阵 H 的 Moore-Penrose 广义逆, T' 是 T 的转置;解 $\hat{\beta}$ 可使训练误差最小,且得到权值的最小范式以及最优泛化性能,并且具有唯一性。

ELM 学习算法过程如下。

(1) 随机设置输入权值 ω_i 以及阈值 $b_i, i = 1, \dots, L$; 其中 L 为隐层节点个数。

(2) 选择一个无限可微的函数作为隐层神经元的激活函数, 并计算隐层输出矩阵 H 。

(3) 计算隐节点与输出节点的连接权值 $\hat{\beta} : \hat{\beta} = H^+ * T'$ 。

ELM 在训练的过程中不需要调整 ω 和 b 的值, 只需根据相应算法来调整 β 值, 便可获得一个全局最优解, 参数选择的过程简单易用, 训练速度提升明显, 且不会如传统神经网络学习算法 (如 BP 算法) 陷入局部最优。

2 样本数据预处理

为了更好地提高 ELM 的效率和精度, 本文采用模糊聚类对所有原始样本进行了数据预处理^[9], 再把处理后的数据作为 ELM 的输入数据进行训练和预测。

模糊聚类数据预处理步骤。

(1) 建立模糊相似矩阵 A 。

建立模糊相似矩阵 A (以绝对值减数法) 公式:

$$a_{ij} = \begin{cases} 1, & i = j; \\ 1 - C \sum_{k=1}^m |x_{ik} - x_{jk}|, & i \neq j. \end{cases}$$

其中, a_{ik} 为第 i 行第 k 列的属性值; a_{jk} 为第 j 行第 k 列的属性值; C 为某常量, 使 $0 \leq a_{ij} \leq 1; i, j = 1, 2,$

$\dots, n$, 其中 m 为样本属性数, n 为样本数。

(2) 求模糊等价矩阵 $t(A)$ 。

通过褶积求 A 的传递闭包, 即 A 自乘得 $A * A = A^2$, 再自乘, 直到 $A^k = A^{2k}$, 则模糊等价矩阵 $t(A) = A^k = A^{2k}, k \in N$ 。模糊等价矩阵的数据作为 ELM 的样本数据。

3 实例仿真

本文以 21 年二代玉米螟危害数据^[10]为例对虫害进行预测。把第 1 ~ 16 年的数据作为 ELM 学习样本, 其余数据作为预测样本。

影响第二代玉米螟有虫株率 $V(\%)$ 的主要预报要素有^[10]: $S1$ 为春、夏玉米种植面积之比, $S2$ 为第一代残虫量 (头/百株), $S3$ 为 7 月上旬温雨系数, $S4$ 为 7 月上旬雨量 (mm), $S5$ 为 7 月上旬平均温度 ($^{\circ}\text{C}$)。

3.1 数据预处理

预报要素划分为 5 级, 预报对象 V 也划分为 5 级: 1 级, 小发生; 2 级, 中等偏轻发生; 3 级, 中等发生; 4 级, 中等偏重发生; 5 级, 大发生; 如表 1 所示^[10]。按表 1 的预报指标分级标准把 21 年的原始预报指标观测数据化为级值, 如表 2 所示^[10]。

本文中属性个数 $m = 5, i, j = 1, 2, \dots, 21, C = 0.7$, 计算模糊关系矩阵 A 矩阵和模糊等价关系矩阵 $t(A)$ 。得到的模糊等价关系矩阵 $t(A)$ 如表 3 所示。

表 1 预报要素分级

级别	S1	S2	S3	S4	S5	V
1	<0.02	<32	<1.7	<7	<23.4	<10
2	0.02~0.14	32~49	1.7~2.3	7~56	23.4~25.5	10.0~20
3	0.15~8.66	50~97	2.4~4.7	57~99	25.6~26.8	20.1~40
4	8.67~9.43	98~152	4.8~5.5	100~134	26.9~29.1	40.1~60
5	>9.43	>152	>5.5	>134	>29.1	>60

表 2 原始数据分级值

预报要素	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
S1	5	5	4	5	4	3	3	4	3	3	2	3	3	3	2	2	2	2	2	2	2
S2	4	4	1	2	1	2	3	2	2	2	1	3	3	5	1	3	2	3	3	3	2
S3	5	1	3	2	1	3	5	3	2	5	1	2	4	4	3	5	2	5	1	1	4
S4	5	2	4	2	1	3	5	3	2	5	2	3	5	4	3	5	2	5	2	2	5
S5	2	5	3	2	1	3	2	3	2	2	1	4	5	1	3	2	3	2	2	2	3
V	5	4	2	3	1	3	3	3	3	3	2	3	3	4	2	3	2	3	2	2	2

表 3 模糊等价矩阵 t(A)

1	0.49	0.32	0.5	0.49	0.48	0.33	0.5	0.49	0.5	0.5	0.5	0.4	0.5	0.5	0.49	0.5	0.49	0.5	0.5	0.5
0.49	1	0.32	0.49	0.49	0.48	0.33	0.49	0.91	0.49	0.49	0.49	0.4	0.49	0.49	0.74	0.49	0.74	0.49	0.49	0.49
0.32	0.32	1	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32	0.32
0.5	0.49	0.32	1	0.49	0.48	0.33	0.58	0.49	0.59	0.58	0.53	0.4	0.58	0.58	0.49	0.58	0.49	0.59	0.59	0.58
0.49	0.49	0.32	0.49	1	0.48	0.33	0.49	0.49	0.49	0.49	0.49	0.4	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49
0.48	0.48	0.32	0.48	0.48	1	0.33	0.48	0.48	0.48	0.48	0.48	0.4	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48
0.33	0.33	0.32	0.33	0.33	0.33	1	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33	0.33
0.5	0.49	0.32	0.58	0.49	0.48	0.33	1	0.49	0.58	0.6	0.53	0.4	0.74	0.6	0.49	0.6	0.49	0.58	0.58	0.6
0.49	0.91	0.32	0.49	0.49	0.48	0.33	0.49	1	0.49	0.49	0.49	0.4	0.49	0.49	0.74	0.49	0.74	0.49	0.49	0.49
0.5	0.49	0.32	0.59	0.49	0.48	0.33	0.58	0.49	1	0.58	0.53	0.4	0.58	0.58	0.49	0.58	0.49	0.62	0.62	0.58
0.5	0.49	0.32	0.58	0.49	0.48	0.33	0.6	0.49	0.58	1	0.53	0.4	0.6	0.61	0.49	0.61	0.49	0.58	0.58	0.61
0.5	0.49	0.32	0.53	0.49	0.48	0.33	0.53	0.49	0.53	0.53	1	0.4	0.53	0.53	0.49	0.53	0.49	0.53	0.53	0.53
0.4	0.4	0.32	0.4	0.4	0.4	0.33	0.4	0.4	0.4	0.4	0.4	1	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
0.5	0.49	0.32	0.58	0.49	0.48	0.33	0.74	0.49	0.58	0.6	0.53	0.4	1	0.6	0.49	0.6	0.49	0.58	0.58	0.6
0.5	0.49	0.32	0.58	0.49	0.48	0.33	0.6	0.49	0.58	0.61	0.53	0.4	0.6	1	0.49	0.69	0.49	0.58	0.58	0.72
0.49	0.74	0.32	0.49	0.49	0.48	0.33	0.49	0.74	0.49	0.49	0.49	0.4	0.49	0.49	1	0.49	1	0.49	0.49	0.49
0.5	0.49	0.32	0.58	0.49	0.48	0.33	0.6	0.49	0.58	0.61	0.53	0.4	0.6	0.69	0.49	1	0.49	0.58	0.58	0.69
0.49	0.74	0.32	0.49	0.49	0.48	0.33	0.49	0.74	0.49	0.49	0.49	0.4	0.49	0.49	1	0.49	1	0.49	0.49	0.49
0.5	0.49	0.32	0.59	0.49	0.48	0.33	0.58	0.49	0.62	0.58	0.53	0.4	0.58	0.58	0.49	0.58	0.49	1	1	0.58
0.5	0.49	0.32	0.59	0.49	0.48	0.33	0.58	0.49	0.62	0.58	0.53	0.4	0.58	0.58	0.49	0.58	0.49	1	1	0.58
0.5	0.49	0.32	0.58	0.49	0.48	0.33	0.6	0.49	0.58	0.61	0.53	0.4	0.6	0.72	0.49	0.69	0.49	0.58	0.58	1

3.2 仿真结果及分析

选择隐层神经元个数为 14,隐层神经元的激活函数为 sig 函数;由于虫害发生程度分为 1~5 级,对应的目标输出模式分别为 (00001)、(00010)、(00100)、(01000)、(10000),故输出层神经元个数为 5,采用 ELM 预测第 17~21 年的虫害程度。用 ELM 模型得到的目标输出中,对于每个目标输出,当 a 是最大值时,取 $a = 1$,否则 $a = 0$ 。ELM 算法训练结果与实测值的对比如图 1 所示。回测率和预测率均达到了 100%。预测值与模拟结果对应如表 4 所示。

选择隐层神经元个数不同,得到的结果也不相同,在 12~25 之间,回测率一般达到 97%~100%,预测率达到 94%~100%,隐层神经元个数太多都会使预测率呈逐步下降趋势,隐层神经元个数太少,回测率和预测率不高。在学习速度上,本文的实验基于 3.0 GHz 的双核 Intel 处理器、2 GB 内存的 PC 机,实验环境为 MATLAB 2009a 版本,运行预测模型所花的时间为 3.533 5 s,满足了虫情预测实时性要求。

使用文献[3]的 LM 算法,直接采用分级后数据

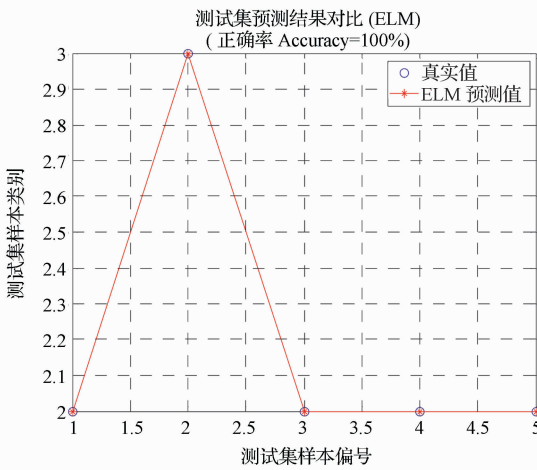


图 1 ELM 算法训练结果与实测值的对比

作样本,在同等条件下进行训练。预测值与模拟结果对应如表 4 所示。同样当 a 是最大值时,取 $a = 1$,否则 $a = 0$ 。模拟结果显示,第二、第三、第四、第五样本预测正确,第一样本预测错误,且第三、第四样本的模拟结果均为 (0.000, 0.999, 0.000, 1, 0.000),0.999 与 1 非常接近,很难判断。采用本文的方法预测精确率明显高于 LM 方法的预测精确率。

表 4 预测值和模拟结果的对应

预测样本	预测值					LM 算法模拟结果					ELM 方法模拟结果				
1	0	0	0	1	0	0.000	0.000	0.871	0.119	0.000	0.851 979	0.268	-1.391	1.839	-0.567
2	0	0	1	0	0	0.000	0.000	0.999	0.001	0.000	0.108 906	-0.007 81	0.747	0.231	-0.079
3	0	0	0	1	0	0.000	0.999	0.000	1.000	0.000	0.623 575	0.200 213	-1.074	1.316	-0.066
4	0	0	0	1	0	0.000	0.999	0.000	1.000	0.000	0.623 575	0.200 213	-1.074	1.316	-0.066
5	0	0	0	1	0	0.000	0.000	0.000	1.000	0.000	0.037 379	-0.017 63	0.3124	0.821	-0.152

4 结语

通过仿真实验可以看出：ELM 在参数选择及学习速度上具有明显的优势,能够满足虫害预测对准确率和实时性的要求,在虫害预测上有广阔的应用前景。下一步可以考虑对 ELM 的方法深入研究,如把结构风险最小等思路与 ELM 结合,进一步提高虫害预测的准确率。

参 考 文 献

1 唐建军,王映龙,彭莹琼. BP 神经网络在水稻病虫害诊断中的应用研究. 安徽农业科学,2010;38(1):199—200,204

2 陈利星,陈绘画,周钦富. 神经网络对马尾松蛀干类害虫数量的混沌识别. 安徽农业科学,2011; 39(28),17281—17282

3 高艳萍,于 红,崔新忠,等. 基于 L-M 优化算法的水稻螟虫预测

模型及其初步应用. 农业工程学报,2007;23(7):360—362,561

4 Huang G B, Zhu Q Y, SIEW C K. Extreme Learning Machine: Theory and Applications. Neurocomputing, 2006; 70 : 489 - 501

5 黄宴委,吴登国,李 竣. 基于极限学习机的结构健康监测数据恢复. 计算机工程,2011;37(16):241—243

6 陈盛双. 基于极限学习机的 XML 文档分类. 计算机工程,2011; 37(19):177—178,182

7 潘华贤,程国建,蔡 磊. 极限学习机与支持向量机在储层渗透率预测中的对比研究. 计算机工程与科学,2010;(2):131—133

8 丁 姣,蔡建荣,张海东,等. 近红外结合 Si-ELM 检测食醋品质指标. 食品与机械,2012;28(1):93—96

9 原思聪,刘道华,刘宏兵,等. 基于模糊预处理的聚类神经网络模型设计. 西安建筑科技大学学报(自然科学版),2005;37(2): 274—277

10 周潘金,张云山. 用 Fuzzy 综合评判二代玉米螟有虫株率. 昆虫知识,1990;27(3):129—132

Applied Research of ELM in the Pest Forecast

WEI Yan-ling

(Department of Electronic and Information Engineering, Liu Zhou Vocational & Technical College, Liuzhou 545006, P. R. China)

[Abstract] ELM (Extreme Learning Machine) is easy to use. It has fast learning speed and good generalization. pests forecast based on ELM are researched. In order to improve the efficiency and accuracy of ELM , ELM is used to pretreat all samples and then trained, forecasts the processed data as the input data of the ELM. Simulation Experimental results show that ELM has high accuracy and good generalization which meets the requirements about accuracy and real-time. A new method is supplied to forecast pests according to the instance.

[Key words] extreme learning machine fuzzy clustering pests forecast