

误差反向传播神经网络模型的改进及其应用

李宗坤¹, 郑晶星², 周晶¹

(1. 大连理工大学 海岸近海工程国家重点实验室, 辽宁 大连 116024; 2. 广东省水利水电勘测设计研究院, 广东 广州 510170)

摘要: 针对误差反向传播神经网络模型(BP模型)学习收敛速度慢、容易陷入局部极小点等缺点, 提出原始数据的非线性规格化函数、记忆式初始权值、阈值和参数自动优选等方法, 对BP模型进行改进。结合陆浑水库大坝沉降实测资料的分析, 编制了改进后的BP网络算法程序。分析结果表明, 改进后的BP模型对于提高网络运行收敛速度、防止陷入局部极小点、克服手工调试参数的盲目性、提高模型精度等都有较大的改善作用。

关键词: BP模型; 非线性规格化; 记忆式初始权值; 阈值

中图分类号: TV698 **文献标识码:** A

收稿日期: 2002-07-01

基金项目: 河南省青年骨干教师资助项目

作者简介: 李宗坤(1961-), 男, 河南邓州人, 郑州大学副教授, 大连理工大学博士研究生, 主要从事大型水工结构安全性评价和结构计算。

BP模型是神经网络模型中最具代表性和广泛应用的一种, 其结构简单, 可操作性强, 能模拟任意的非线性输入、输出关系。但它也存在很多缺点, 如学习收敛速度慢, 通常需要经过几千次迭代甚至更多, 容易陷入局部极小点而无法得到全局最优解。目前已有不少人对此提出了改进的方案。如文献^[1]提出在修改权值中加入“动量项”; 文献^[2]采用共轭梯度法; 文献^[3]引入同伦方法; 文献^[4]中采用Catchy误差估计器代替传统的LMS误差估计器; 文献^[5]将输入输出数据初始化到(0.1, 0.9)的区间内。这些方案对BP模型起到一定的改进作用, 但仍然存在不足。作者通过对大量实例的分析研究, 提出了一些新的改进方法。

1 BP模型的改进

1.1 原始数据的非线性规格化 大多数BP模型以S形函数作为转换函数, 该函数的值域为[0, 1], 因此在训练时要将原始数据规范到[0, 1]区间, 通常采用标准的归一方法来实现。但是规范后的每个输出的教师值序列中至少有一个值为0, 一个值为1, 恰好是S形函数的极小值和极大值, 要求联接权足够大才能使网络的输出值与其匹配, 从而需要相当多的训练次数来不断修正权值, 导致训练速度缓慢。为避免这种现象, 一些文献建议将教师值进行线性规范化至某一区间内^[5,6], 如文献^[6]提出按

$$y = y / (y_{\max} + y_{\min}) \quad (1)$$

进行处理。这样做可以加快网络学习速度, 但是, 在进行预报时, 由于S形函数的值域为[0, 1], 因而预报值 $y < 1$, 即 $y / (y_{\max} + y_{\min}) < 1$, 反规格化后得 $y < y_{\max} + y_{\min}$ 。可知, 预报值 y 总是小于或等于原数列中的最大值与最小值之和, 当 $y_{\min} = 0$ 时, $y < y_{\max}$, 即预报值 y 将不可能超过原数列中的最大值, 这无疑将预报值限定在某一数值范围内, 大大限制了BP模型预测能力的发挥。本文建议采用非线性规格化数据的方法, 特别是针对坝体沉降的特点, 提出采用函数

$$y' = 1 - \frac{1}{e^{[0.233+1.386*(y-y_{\min})/(y_{\max}-y_{\min})]}} \quad (2)$$

来预处理数据。实践表明,采用该方法规格化后的数据不仅预报效果更好,且在提高拟合精度方面也有一定改善作用。

1.2 记忆式初始权值、阈值 收敛速度慢、迭代时间长是BP算法的主要问题之一。通过大量实践发现,BP算法的初始权值、阈值虽然在理论上可以随意选取,但其选取方法却对BP算法的收敛速度有很大影响。一些文献也提出了相应的初始权值、阈值选取方法^[5],他们在各自的研究领域内取得了一定的成效。本文在此基础上提出记忆式初始权值、阈值方法。

$$\begin{aligned} w_1^0(\cdot) &= Rnd(\cdot), \theta_1^0(\cdot) = Rnd(\cdot) \\ w_i^0(\cdot) &= w_{i-1}^1(\cdot), \theta_i^0(\cdot) = \theta_{i-1}^0(\cdot) \quad (i = 2, 3, \dots) \end{aligned} \quad (3)$$

式中: $w_i^0(\cdot)$, $\theta_i^0(\cdot)$ 分别为第*i*次网络运行时的初始权值、阈值; w_i^1 , θ_i^0 分别为第*i*次网络运行结束时权值、阈值。实践证明,采用记忆式初始权值、阈值方法,对于网络运行收敛是高效的,特别是在大量的调试工作中,同时对于防止陷入局部极小点也有积极的作用。

1.3 参数自动优选 影响BP模型建模效果的参数较多,如隐含层的单元数、学习率、惯性冲量*u*等。这些参数的选取并无确定公式可循。且随着所研究问题的不同参数值也变动较大,因而研究人员常常需要经过繁杂的手工调试及大量的对比分析后才能得出一组最理想的参数值。本文采用参数的循环嵌套方法,同时以确定性系数^[7]最大作为优选的评价标准,从而实现参数的自动优选。

$$D = 1 - \frac{S^2}{\sigma^2} = 1 - \frac{\sum_{j=1}^N (y_j - \hat{y}_j)^2}{\sum_{j=1}^N (y_j - \bar{y}_j)^2} \quad (4)$$

式中: D 为确定性系数; y_j 为实测值; \hat{y}_j 为计算值; \bar{y}_j 为实测序列均值; N 为建模样本数。采用参数自动优选方法,可以避免手工参数调试的盲目性,由计算机取代研究人员的繁重调试工作,而直接获取最终的分析结果。

2 实例分析

大坝安全监测的主要目的是为了了解坝体及基础的实际状态,尽早发现异常现象,进行必要的处理,以确保大坝的安全运行。位移是监测的重要物理量,坝体在各种荷载作用下的位移分布规律是判断坝体结构是否正常的重要依据,也是用来预报和预测的手段之一。近年来,人工神经网络理论及应用得到飞速发展。由于人工神经网络BP模型所反映的函数关系不必用显式的函数表达式表示,而通过调整网络本身的权值和阈值来适应,可避免因为因子选择不当而造成误差。因而,利用BP模型对大坝原型观测资料进行建模分析已成为一个新的有重要意义的研究课题。

2.1 工程概况 陆浑水库位于黄河支流伊河上,1965年建成蓄水,最大坝高55m,库容13.2亿 m^3 ,坝型为粘土斜墙(加截水槽)坝。陆浑水库管理处,自1996年起开始对大坝进行位移观测和渗流观测工作,积累了30多年系统观测资料,为分析工作奠定了基础。

2.2 网络的结构 对于土石坝原型观测资料的 BP 模型网络,其输入层结点数一般取影响因子数,输出层结点数 1 个,即效应量。根据坝工理论及陆浑水库大坝观测资料的初步分析可知,影响大坝沉降的主要因素为时效及水压(上下游水位差),同时考虑到坝地质条件的复杂多变性及不可量化性,拟定采用历时、水压及上一次沉降观测值作为输入层的 3 个单元(方案 A),同时分别采用历时与水压、历时与上一次沉降观测值作为输入层单元(方案 B、方案 C),并将这 3 种方案的建模效果进行了对比。

2.3 参数的选取 为避免参数手工调试的繁琐,采用参数自动优选方法,主要对隐层单元数、学习率、惯性冲量 u 进行组合优选。目前对隐层单元数取值范围的研究较多,但并没有定论^[6]。本例采用 1 到 3 倍输入层单元数即 1 到 9 作为隐层单元数的取值区间,而学习率、惯性冲量 u 均取在(0, 1)。另外,训练精度的取值可参考文献^[6]。

2.4 实例计算 根据 BP 模型建模步骤,编制了改进后的 BP 网络算法程序。对陆浑水库大坝 P4 测点 1977 年至 1995 年的沉降(表 1)建立了 BP 网络模型,并对其 1996 年及 1997 年沉降值进行预测。各计算成果如表 2~表 5 所示。

 表 1 P₄测点 1977 年至 1997 年间各历时、

水压、沉降测值

序号	日期	历时/d	水压/m	累计沉降/mm
1	1977 年 5 月	4332	26.94	43.70
2	1978 年 5 月	4697	23.64	44.50
3	1979 年 5 月	5062	25.19	47.70
4	1980 年 5 月	5428	28.60	52.30
5	1981 年 5 月	5793	28.74	54.21
6	1982 年 5 月	6158	29.33	55.58
7	1983 年 5 月	6523	32.92	55.70
8	1984 年 5 月	6889	31.87	57.70
9	1985 年 5 月	7254	33.07	58.60
10	1986 年 5 月	7619	29.36	60.24
11	1987 年 5 月	7984	27.96	59.13
12	1988 年 5 月	8350	30.18	57.00
13	1989 年 5 月	8715	37.84	57.30
14	1990 年 5 月	9080	38.86	59.00
15	1991 年 5 月	9445	35.18	60.20
16	1992 年 5 月	9811	28.81	61.80
17	1993 年 5 月	10176	34.57	63.17
18	1994 年 5 月	10541	37.49	66.23
19	1995 年 5 月	10906	29.30	61.80
20	1996 年 5 月	11272	32.47	62.03
21	1997 年 5 月	11637	38.24	65.30

注:水压为上下游水位差。

表 2 BP 模型与逐步回归模型建模精度对比

残差区间	样本所占比例数(%)	
	BP 模型	逐步回归模型
(0, 1)	47.1	23.5
(1, 3)	41.2	52.9
(3, 5)	17.6	17.6
(5,)	5.9	17.6

表 3 不同初始权值、阈值方法计算成果对比

	总迭代次数	总耗时	陷入局部极小点总次数	最终模型总偏差/mm ²
记忆式初始权值阈值	5542	4	0	34.34
(0, 1)随机初始权值阈值	866681	9 06	199	34.34

$$\text{注: 最终模型总偏差 } E = \sum_{j=1}^N (\hat{y}_j - y_j)^2$$

表 4 非线性与线性规格化函数计算成果对比

规格化函数	平均拟合误差(%)	平均预报误差(%)
非线性公式(2)	1.91	2.03
线性公式(1)	5.38	2.59

$$\text{注: 平均拟合误差 } E = \frac{1}{N} \sum_{j=1}^N |100 \times (\hat{y}_j - y_j) / y_j|$$

$$\text{平均预报误差 } E = \frac{1}{m} \sum_{i=1}^m |100 \times (\hat{y}_j - y_j) / y_j|; m \text{ 为预报样本数。}$$

表 5 输入层单元数 3 种方案计算成果对比

方案	输入层单元数	平均拟合误差(%)	平均预报误差(%)

3 结 语

(1)由 BP 模型与逐步回归模型建模精度对

比(表 2)可以看出, BP 模型的建模效果明显优于逐步回归模型。这说明人工神经网络 BP 模型本身具有较好的自适应性和鲁棒性, 更能反映事物的非线性内在规律。(2)不同初始权值、阈值方法计算成果(表 3)表明, 在运行结果相同

			(%)	差(%)
A	$3(a+b+c)$		1.91	2.03
B	$B2(a+b)$		2.01	2.22
C	$C2(a+c)$		2.09	3.81

注: a、b、c 均指输入层单元, 其中 a 指历时; b 指水压;
c 指上一次的沉降测值

的情况下, 采用记忆式初始权值、阈值方法与采用(0, 1)随机初始权值、阈值方法, 其总运行迭代次数相差甚大, 前者只有后者的 1/156, 因而前者耗时很少而后者耗时很多。同时, 前者从未陷入局部极小点, 而后者却达 199 次之多。这充分说明了记忆式初始权值、阈值方法对于提高多次 BP 网络运行时的收敛速度有着重大的实用价值, 对于防止陷入局部极小点也有很好的效果。(3)由非线性与线性规格化函数计算成果(表 4)可以看出, 采用非线性规格化函数是合理的, 尤其是对于本身具有递增规律的数列, 由于非线性规格化函数没有人为限制数列的最大界值, 因而更能反映该数列的本质规律。(4)通过输入层单元数 3 种方案计算成果的对比如(表 5)分析, 可以看出, 采用历时、水压及上一次沉降测值作为输入层的 3 个单元是比较合理的。这也体现了人工神经网络 BP 模型在复杂因子下建模的优越性。

综上所述, 采用非线性规格化函数对数列进行预处理, 消除了线性规格化函数处理数列时对预报值范围的限制; 采用记忆式初始化权值、阈值方法, 大大提高了多次 BP 网络运行时的收敛速度; 采用参数自动优选方法建立 BP 模型, 避免了手工调试参数的繁琐工作。

参考文献:

- [1] Rumelhart D E, et al. Learning Representation by BP Errors [J]. Nature(London), 1986, 7.
- [2] Patrick P, et al. Minimization Method for Training Feed Forward Neural Network [J]. Neural Network, 1994, 7.
- [3] 高小榕, 杨福生. 采用同伦 BP 算法进行多层前向神经网络的训练 [J]. 计算机学报, 1996, (9): 687-694.
- [4] 彭松, 方祖祥. BP 神经网络学习算法的联合优化 [J]. 电路与系统学报, 2000, (7): 26-30.
- [5] 赵斌, 吴中如, 等. BP 模型在大坝安全监测预报中的应用 [J]. 大坝观测与土工测试, 1999, (12): 1-4.
- [6] 刘国东, 丁晶. BP 网络用于水文预测的几个问题探讨 [J]. 水利学报, 1999, (1): 65-70.
- [7] 冯国章, 李佩成. 人工神经网络结构对径流预报精度的影响分析 [J]. 自然资源学报, 1998, (4): 169-174.