

文章编号: 1005-2542(2007) 05-0563-05

考虑问题随机性的神经网络

潘 华¹, 李本强¹, 雷元欣²

(1. 五邑大学, 江门 529020; 2. 佛山科学技术学院, 佛山 528000)

【摘要】

关键词: ; ;
中图分类号: TU 4 文献标识码:

Neural Network Study on Random Problem

PAN Hua¹, LI Ben-qiang¹, LEI Yuan-xin²

(1. Wuyi Univ., Jiangmen 529020, China; 2. Foshan Univ., Foshan 528000, China)

【Abstract】 The study on artificial neural network dealing with random samples is a challenging problem. For an artificial neural network model, the choice of the error of training and the evaluation of the predicted results is concerned. To our knowledge, the error of training is mostly chosen by trial-and-error. The predicted results were often evaluated by using relative error method, and it is often regarded that the less the relative error is, the better the predicting precision is. But these methods are no longer fitted for random training samples. Therefore, combined with the design of concrete mix proportion, by taking into account the stochastic characteristic of training samples, a new method for determining training error and evaluating predicted results of a neural network model was studied. Several formulas for random training samples were provided. In the end, a practical example was given. Experimental results justify that the method studied in this paper is useful, and its conclusion is close to engineering actualities. It has certain practical significance.

Key words: neural network; prediction; randomness

神经网络的研究是从人脑的生理结构出发来研究人的智能行为, 模拟人脑信息处理的功能。它具有许多优良性能, 如允许大规模并行处理、高度的非线性映射能力和泛化能力等, 适合于在特定的

输入和输出之间建立复杂的映射关系, 这对于那些有一定的实际经验却又难于建立有效数学力学模型的工程领域具有重要的现实意义, 这方面的研究工作非常丰富, 如: Shi^[1] 用累积分布函数法处理神经网络的样本数据, 与线性变换方法相比, 大大降低了网络模型预测的错误率。文献 [2] 中提出了多变量神经自适应控制器通用结构。文献 [3, 4] 中分别给出了网络结构每层隐节点数的起始值和上限值的计

收稿日期: 2006-11-14 修订日期: 2007-03-28

作者简介: 潘 华(1963), 女, 硕士。研究方向为土木工程及系统工程在土木工程中的应用。lb_ph@126.com

Sel	建立了钢筋
二献	究了训练样
六,提	的计算式。
的基	型的模糊
文南	出了利用
能区	蒙特卡
网络	地区人
献	了
所	段。文
基	应用
网络	部
隆宽	,经
13,	经网
中	上
16	了
型	元
住	作
神	来讨
21	网
23	神经

可 尽管以
解决的 题有所不
神经网络 输入输出
虑实际 变量的限
性问题 存在于各
混凝土 比问题,由
合比 一般作为网
混凝土 强度(网络输
神经网络 说,就会遇到
方面的 难。对于确定
种据实际 问题的要求

目本 身 认为相对误差超
精度 1% 随机问题,训练误差的确定不
的简单 经过反复试算调整确定¹²,带有
的盲 而神经网络训练误差的大小直
系到训 时间的长 训练时间
长,甚至训练瘫痪;取
对于网 结果
不合适了
是预测
机性的
实际 将以
在 土木工程
随机性

$Y_i = a_2$
在工程应用中
混凝土取3种
值,即上

结果的评价等问题,都将与输入输出变量为确定性关系时有所不同。

2 网络训练误差的选取

众所周知,学习算法中的关键是在学习过程中以尽可能快地减小训练误差的方式进行,通常是梯度方向。但减小到什么程度学习结束呢,即如何选取训练误差控制值,这是在实际应用中经常碰到的问题,一般是经过反复试算调整确定,带有一定的盲目性。一般地,神经网络模型的输入输出变量按其性质的不同可分为确定性变量和非确定性变量(本文涉及的是随机变量)。对前者相对容易些,可以结合实际对误差的要求计算确定,而对后者为非确定性变量就更为复杂。下面结合混凝土性能指标的随机特性,研究神经网络模型的输出变量为随机变量时确定训练误差控制值的方法。

由第 1 节中分析知,作为训练样本的 Y_i 不一定是混凝土性能指标 μ_i 的真实反映,训练误差的大小反映的是网络预测计算值 μ_i 与实测值 Y_i (而非 μ_i) 之间的差值大小,这样训练误差 E 的大小并不一定能反映训练结果的好坏。但通过式(2)可以得到 μ_i 的置信区间,因此,若通过训练使网络预测计算值落在该置信区间内,就可以认为此时的样本学习反映了 μ_i 的情况,学习就可以结束。因此,设某一性能指标 μ_i 的置信区间的 $1/2$ 为 Δ ,

$$\mu_i - Y_i \leq \Delta \tag{3}$$

则全网络平均训练误差

$$E = \frac{1}{2} \sum_{i=1}^n (\mu_i - Y_i)^2 \leq \frac{1}{2} \sum_{i=1}^n \Delta^2$$

即

$$E \leq \frac{1}{2n} \sum_{i=1}^n \Delta^2 \tag{4}$$

式中 n 为计算 Y_i 时所采用试件的个数,为网络输出变量总数。

由于激励函数 Sigmoid 的函数值输出范围在接近 0,1 时,曲线较平缓,变化速度非常慢,为了减少网络的学习时间,将输入输出数据归一化变换为 0.1,0.9 之间⁶,这样 Sigmoid 函数在该区间内变化梯度比较大,网络收敛时间大大缩短,改善了网络性能,其变换方法为

$$= 0.1 + \frac{0.8(\mu - \min)}{\mu_{\max} - \min}, \quad \mu \in [0.1, 0.9] \tag{5}$$

式中 μ 为样本原始值, μ_{\max} 、 μ_{\min} 是样本原始值中的最大、最小值,为归一化后的值。

由式(5)可得 Y_i 经归一化处理后的均方差为

$$\sigma_i = \frac{0.8}{Y_{i,\max} - Y_{i,\min}} \sigma \tag{6}$$

将式(6)代入式(4),得

$$E \leq \frac{0.32}{n} \sum_{i=1}^n \left[\frac{\sigma_i}{Y_{i,\max} - Y_{i,\min}} \right]^2 \tag{7}$$

因此,全网络平均初始训练误差取为

$$\varepsilon = \frac{0.32}{n} \sum_{i=1}^n \left[\frac{\sigma_i}{Y_{i,\max} - Y_{i,\min}} \right]^2 \tag{8}$$

3 预测结果的评价方法

对模型预测的结果,一般是用平均相对误差值(如 5)来评价^{12,19}。由于混凝土的性能指标具有随机性,即用来训练的样本本身是随机变量,显然此时相对误差值的大小不能反映预测结果的好坏。因此,下面将针对混凝土性能指标的随机性这一特点,研究合适的评价方法。

由第 1 节的分析知,作为训练样本的 Y_i 不一定是混凝土性能指标 μ_i 的真实反映,因此,预测值与实测的 Y_i 的接近程度不一定能反映预测的优劣。实际上,对预测结果的评价面临着两方面的问题:①首先要判断出每次预测的正确性;②每次预测正确的概率如何。现分析如下

由式(2)可得平均数 μ 的置信区间,即

$$\left[Y_i - \alpha \frac{\sigma_i}{\sqrt{n}}, Y_i + \alpha \frac{\sigma_i}{\sqrt{n}} \right]$$

因此可以认为,对于神经网络的每次预测,若神经网络预测的结果落在该置信区间内(或预测值 μ_i 与 Y_i 之差的绝对值 $\leq \alpha \frac{\sigma_i}{\sqrt{n}}$),则接受该预测结果,预测正确;否则,拒绝它,预测失败。其可信度为 $1 - \alpha$ (如 α 取 0.05,则可信度为 95%)。

因为一次预测正确与否并不能正确反映模型的预测性能,必需考查多次预测的情况。因此,定义如下预测正确率

$$\text{正确率}(\%) = \frac{\text{测试正确的次数}(n)}{\text{所有的测试次数}(N)} \tag{9}$$

显然正确率的大小能够反映模型的预测性能。然而由于混凝土性能指标的随机性,在有限的测试次数 N 的条件下,该正确率的可信度如何呢?对此,可以先通过式(9)计算预测正确概率的估计值,然后再分析出预测正确的概率的置信区间。

由于每次的预测,要么正确,要么失败。因此,预测的结果可以用(0-1)分布的随机变量来描述。设随机变量 X 表示模型一次预测的结果,每次预测

正确的概率为 α ，则 λ 的分布为

$$P(\lambda = \lambda_0) = (1 - \alpha)^{\lambda_0 - 1}, \quad \lambda_0 = 0, 1 \quad (10)$$

由于 λ 未知, 为估计 λ , 选取充分大容量的检测样本 x_1, x_2, \dots, x_N , 这样由中心极限定理

$$\begin{aligned} &= \frac{x_1 + x_2 + \dots + x_N - N}{\sqrt{N} \sqrt{(1 - \alpha)}} = \\ &= \frac{N\lambda - N}{\sqrt{N} \sqrt{(1 - \alpha)}} = \frac{N(\lambda - 1)}{\sqrt{N} \sqrt{(1 - \alpha)}} \quad (11) \end{aligned}$$

近似服从 $N(0, 1)$ 分布。于是有

$$P\left\{-\alpha_2 \leq \frac{N(\lambda - 1)}{\sqrt{N} \sqrt{(1 - \alpha)}} \leq \alpha_2\right\} \approx 1 - \alpha \quad (12)$$

由式(12)可得

$$\lambda_1 = 1 - \alpha_2 \sqrt{(1 - \alpha)}, \quad \lambda_2 = 1 + \alpha_2 \sqrt{(1 - \alpha)} \quad (13)$$

式中

$$\begin{aligned} \lambda_1 &= \frac{1}{2a}(-\alpha_2 - \sqrt{\alpha_2^2 - 4a}) \\ \lambda_2 &= \frac{1}{2a}(-\alpha_2 + \sqrt{\alpha_2^2 - 4a}) \end{aligned}$$

其中

$$\begin{aligned} a &= N - \frac{\alpha_2^2}{4}, \quad \alpha_2 = N - \alpha \\ &= - (2N - \frac{\alpha_2^2}{2}) \end{aligned}$$

于是得网络模型预测正确的概率 λ 的置信区间为

$$(\lambda_1, \lambda_2) \quad (14)$$

显然, 预测正确的概率 λ 能反映出模型的预测性能, 比用相对误差方法更科学合理。

4 实例计算

根据当地材料及自密实混凝土的性能要求, 用

正交试验的试验结果作为样本集(共 27 组)。将各种材料的用量选为输入变量, 即将水泥、矿物掺合料、化学外加剂、石子、河砂以及拌和水作为网络的输入变量; 混凝土的各项性能指标选为输出变量。即强度、坍落度和扩展度为网络的输出变量。强度、坍落度、扩展度 3 个指标的方差分别为 5.3, 1.6, 4.2, 计算得, 初始训练误差为 0.018。隐层节点数初始值为 6, 取置信概率为 95%。

为了增加训练和测试样本对, 将已有的 27 组试验, 随机地分成两部分, 如: ①将其中 18 组作为训练样本集, 剩下的 9 组构成测试样本集, 对网络进行训练和测试, 得出测试正确的次数; ②又将这 27 组试验, 随机地分成两部分, 这两部分与 ①不同, 重复步骤 ①; 如此不停地随机将已有的 27 组试验分成两部分, 并重复步骤 ①, 直到测试样本的数量增大到符合要求, 如 100。

由以上测试数据经过网络训练、模型检验、性能预测等环节进行测试、修改。其中部分计算结果见表 1。从表中可以看出, 由神经网络模型得出的预测正确率较高, 满足工程要求。

同样地, 仍然采用与以上相同的试验数据进行网络训练、模型检验等, 要达到表 1 的预测精度, 如果没有用式(8)计算的初始训练误差, 而是随机地取一个值, 如果该值较大, 此时训练时间虽然短些, 但达不到表 1 同样的预测精度; 若随机值比较小, 则要经过很长时间的训练, 甚至可能网络瘫痪。有了式(8)使误差值的选取有了理论依据。

表 1 实测/计算结果分析

序号	kg	kg	kg	kg	A	kg	强度 M a	坍落度 mm	扩展度 mm
							(实测 计算)	(实测 计算)	(实测 计算)
1	265	135	824	1 006	8.8	170	51.0 51.3	22.0 22.2	59 58.3
2	265	170	870	979	7.4	185	52.2 53.2	23.5 21.8	60 61.9
3	265	180	857	928	7.56	170	54.6 60.6	25.5 23.4	65 60.4
4	370	135	701	1 009	8.59	185	61.5 58.2	23.0 22.1	62 62.8
5	370	170	760	930	9.18	170	73.1 69.9	24.5 24.2	63 62.7
6	370	170	804	871	11.8	185	67.5 68.9	25.5 24.4	68 66.8
7	420	180	734	896	16.8	170	70.2 73.8	25.0 25.3	68 67.1
8	420	135	747	913	12.21	185	69.3 68.0	25.0 24.1	66 66.1
9	420	180	676	973	13	150	76.2 75.2	25.5 25.1	66 61.5
预测正确率(预测正确次数/预测总数)							97	91	89
预测正确的概率的置信区间							(92, 99)	(84, 95)	(81, 94)

注: 表中黑体数据表示该次预测超过置信区被拒绝

结 语

本文在分析了高维非线性问题的复杂性基础上,应用遗传算法对神经网络训练差初始值的计算进行了有效的求解方法。从而减小了初始值对神经网络训练效率的影响,提高了神经网络正确率这一概念。进一步由神经网络预测正确的概率,对神经网络评价的可靠性。应用结果表明该方法是有效的。结论更贴近工程实际。且为推广应用神经网络工程应用提供理论依据,具有一定的工程实用价值。

值得注意的是一旦对随机样本、训练误差选取不当,会造成神经网络瘫痪。

参考文献:

- 1 Shi J J, Hammer, S. Improving prediction accuracy by transforming input data for neural networks. Journal of computational engineering, 1990, 10(10): 109-116.
- 2 金耀初. 神经网络模糊推理量测模糊推理响应速度. 模糊数学, 1994, 4(1): 228-231.
- 3 王元, 寇志. 神经网络及其在工程中的应用. M. 北京: 机械工业出版社, 1994, 34-36.
- 4 高大. 具有非线性转移函数前向三层神经网络的研究. 计算机学报, 1998, 21(1): 80-85.
- 5 Seleemh. A neural network model for predicting maximum shear capacity of concrete beams. Journal of bridge engineering, 1997, 2(1): 1-10.