

# BP 网络学习能力与泛化能力之间的定量关系式

李祚泳, 易勇鸷

(成都信息工程学院, 四川成都 610041)

**摘 要:** 分析 BP 网络过拟合时网络学习能力与泛化能力之间的内在联系, 引入描述问题复杂性程度的复相关系数, 建立了 BP 网络过拟合时, 反映网络学习能力的训练样本集的训练相对误差与表征泛化能力的网络对检验样本集的测试相对误差之间满足的定量关系式. 通过模拟若干不同类型函数的 BP 网络数值建模试验, 确定了关系式中过拟合参数  $q$  的取值范围为  $0.007 \sim 0.07$ , 指出 BP 网络应用于给定样本集的训练过程中, 具有较佳泛化能力的停止训练方法.

**关键词:** BP 网络; 学习能力; 泛化能力; 过拟合关系式

**中图分类号:** TP18 **文献标识码:** A **文章编号:** 0372-2112 (2003) 09-1341-04

## Quantitative Relation Between Learning Ability and Generalization Ability of BP Neural Network

LI Zuoyong, YI Yongzhi

(Chengdu University of Information Technology, Chengdu, Sichuan 610041, China)

**Abstract:** Based on the analysis of the internal relation between learning ability and generalization ability of the overfitting of BP neural network, by the introduction of multi-correlative coefficient to manifest the complexity of the function, a quantitative uncertainty relation between the fitting relative error of training sample sets and the testing relative error of verifying sample sets, which describe the learning ability and the generalization ability of BP network, respectively, was revealed in the overfitting of BP neural network. Tests of numerical simulation for multi-kinds of different functions were carried out to determine the value distribution ( $0.007 \sim 0.07$ ) of overfitting parameter  $q$  in the relation. Based on the quantitative relation, the training method for the improvement of generalization ability in the training process of sample sets using BP neural network was given.

**Key words:** BP neural network; learning ability; generalization ability; overfitting relation

### 1 引言

学习能力与泛化能力是 BP 网络性能的两个最重要的考核指标, 它们的高低是评定网络是否适合问题求解的一种很好的尺度. 因此, 研究 BP 网络的学习能力与泛化能力及其它们之间的关系愈来愈受到人们的关注. 网络的泛化能力研究最多的是神经网络结构复杂性和样本复杂性对神经网络泛化能力的影响. 对于线性阈值的单隐层全连接前向网络, Baum 得出了为保证固定结构的神经网络泛化能力所需训练样本数的上界和样本复杂性的下界<sup>[1]</sup>. Moody 研究了泛化误差与训练集误差之间关系, 证实了实值神经网络结构设计最简原则的 Moody 准则<sup>[2]</sup>. Barron 等给出了学习后神经网络的泛化误差同样本数和隐节点数之间的关系, 得出随着隐节点数增加, 逼近误差将逐渐减小, 而估计误差将逐步增大, 好的泛化能力取决于二者的协调<sup>[3,4]</sup>. Cataltepe 等提出当网络的泛化误差达到最小时停止学习, 用以改进神经网络的泛化能力<sup>[5]</sup>. Amari

等研究了交叉测试法 (CV) 中测试样本数占总样本数比例对神经网络泛化能力的影响<sup>[6]</sup>. 我国学者研究了为保证网络的泛化能力所需的样本数<sup>[7]</sup>、网络的泛化能力与结构复杂性和样本复杂性的关系<sup>[8,9]</sup>、基于主要影响因素和修正误差函数增强网络泛化能力的方法<sup>[10]</sup>和提高泛化能力的若干途径与实际方案<sup>[11,12]</sup>. 笔者建立了 BP 网络出现过拟合时, 网络的权值改变量与表征泛化能力的网络对新样本集的辨识误差之间满足的不确定关系式<sup>[13]</sup>. 已有的研究大多集中于研究训练样本复杂性和网络结构的复杂性对网络泛化能力的影响, 而过拟合出现时学习能力与泛化能力之间关系的定量研究则未见报道.

本文在文献[13]的研究基础上, 进一步建立了 BP 网络过拟合时, 代表网络的学习能力的训练样本集的训练相对误差与代表泛化能力的检验样本集的测试相对误差之间应满足的定量关系式, 通过对多种不同类型函数的大量数值模拟试验, 确定出使关系式成立的公式中参数的取值范围, 并根据此关

式指出 BP 网络训练过程中,为使网络有好的泛化能力的最佳停止训练法.

## 2 BP 网络过拟合出现时的泛化能力与学习能力之间满足的关系式

文献[13]指出:BP 网络过拟合出现时,权值改变量  $w$  与未参与训练的样本集的辨识误差  $|y|$  之间应满足过拟合关系式

$$|w| + |y| = \frac{Rhp}{2\log_2(1 + M/N)} \quad (1)$$

式中:  $R$  为所逼近函数的复相关系数;  $h$  为所构造的 BP 网络的隐节点数;  $M$  和  $N$  分别为网络训练完  $t$  遍后,训练样本集的网络输出平均归一化值和输出的方均根误差;  $p$  为满足关系式成立的“过拟合参数”,其取值范围为  $1 \times 10^{-5} \sim 5 \times 10^{-4}$ ;

$w$  为网络训练过程中出现过拟合时,样本集相继训练完两遍后的相对权值改变量;  $y$  为网络对未参与训练的新样本集的辨识误差,  $w$  和  $y$  的计算式详见文献[13].

事实上,随着 BP 网络训练继续进行,权值改变量  $|w|$  逐渐减小,拟合效果愈来愈好,即  $|z/z|$  逐渐减小,而网络对新样本的平均测试相对误差  $|y/y|$  与辨识误差  $|y|$  变化趋势是一致的.因此,若用  $|z/z|$  和  $|y/y|$  分别取代式(1)左端的  $|w|$  和  $|y|$ ,不确定关系式(1)形式仍然成立,不过右端的“过拟合参数” $p$  的取值范围不同,记为  $q$ ,从而下列过拟合不确定关系式成立

$$|z/z| + |y/y| = \frac{Rhq}{2\log_2(1 + M/N)} \quad (2)$$

式中:  $q$  为待定的“过拟合参数”.  $|z/z|$ 、 $|y/y|$ 、 $M$  和  $N$  分别用以下诸式计算:

$$|z/z| = \frac{1}{m} \sum_i \left| \frac{O_i - T_i}{T_i} \right| \quad (3)$$

$$|y/y| = \frac{1}{m} \sum_i \left| \frac{O_i - T_i}{T_i} \right| \quad (4)$$

$$M = \sum_i O_{ki} / m \quad (5)$$

$$N = \sqrt{\sum_i (T_{ki} - O_{ki})^2 / m} \quad (6)$$

## 3 确定过拟合参数 $q$ 的取值的数值模拟试验

本试验设计具有输入层、隐层和输出层的 3 层 BP 网络,并只考虑一个输出节点  $k=1$ . 数值模拟试验过程中,选取输入节点数和隐节点数的变化范围分别为  $i=3 \sim 10$  和  $j=3 \sim 15$ ; 训练样本数和检验样本数变化范围分别为  $m=8 \sim 128$  个和  $m=8 \sim 64$  个,但每次试验的预留检验样本数  $m$  少于训练样本数  $m$ . 一共模拟了如表 1 所示的 5 种不同类型函数数值试验,赋予各类函数中的系数  $a_i$  为  $0 \sim \pm 100$  内的随机数,  $x_i$  为样本的第  $i$  个输入因子值. 对  $y_1$ 、 $y_2$  类函数,  $x_i$  的赋值范围为  $0 \sim \pm 10000/20$  内随机数; 对  $y_3$ 、 $y_4$  和任意类函数,  $x_i$  的赋值范围为  $0 \sim \pm 100/20$  内随机数.

数值试验过程中,采用各类模拟函数定义域内因子随机选取值的归一化值作为网络的因子输入值; 训练样本的期望

输出值是各类模拟函数计算值的归一化值. 确定式(2)中的过拟合参数  $q$  值的数值试验过程与文献[5]中提出的防止网络过度训练过程作法相同: 以  $m$  个学习样本来训练网络, 以  $m$  个检验样本来验证网络; 在网络训练的初始阶段, 网络的训练误差和检验误差将同时减小; 随着网络训练精度的提高, 检验误差的减小可能变慢; 当网络训练精度进一步提高, 检验误差增大时, 停止训练. 计算此时的训练误差  $|z/z|$ 、检验误差  $|y/y|$ 、训练样本集的网络输出平均归一化值和输出的方均根误差之比  $M/N$ . 将  $|z/z|$ 、 $|y/y|$ 、 $M/N$ 、函数的复相关系数  $R$  及网络的隐节点数  $h$  代入式(2), 从而可以确定出  $q$  值. 对上述 5 类函数, 分别进行了数百组不同的过拟合数值试验, 得出一致结果: 当过拟合出现时, 关系式(2)中的过拟合参数  $q$  的取值范围一般为  $q \in [7 \times 10^{-3}, 7 \times 10^{-2}]$ . 5 种不同类型函数的过拟合数值模拟试验部分计算结果见表 1. 关系式(2)可视作 BP 网络过拟合出现时, 学习能力与泛化能力之间应满足的一般关系式. 此关系式说明: 构造一定结构的 3 层 BP 网络, 对某给定样本集进行训练, 当训练进行到一定阶段出现过拟合后, 若训练继续进行, 虽然可以使训练样本集的训练误差  $|z/z|$  进一步减小, 提高了网络学习能力. 但因关系式(2)右端的复相关系数  $R$ 、网络隐节点数  $h$  确定, 过拟合参数  $q$  变化范围也确定, 因子  $\log_2(1 + M/N)$  变化很小. 因此, 可认为关系式右端近似保持不变, 为保持关系式(2)成立, 当左端训练样本集的训练相对误差进一步减小的同时, 检验样本集的测试相对误差必然增大, 即泛化能力降低, 这正是过拟合现象出现的必然结果.

## 4 具有较佳泛化能力的停止训练法

关系式(2)可用于指导 BP 网络训练, 使具有较佳泛化能力. 一般说来, BP 网络训练直到刚开始出现过拟合前, 训练样本集的平均训练相对误差  $|z/z|$  与对检验样本集的平均测试相对误差  $|y/y|$  相差不多, 因此用  $|z/z|$  近似代替  $|y/y|$  后, 关系式(2)也是成立的

$$|z/z|^2 = \frac{Rhq}{2\log_2(1 + M/N)}$$

故有

$$q = |z/z|^2 \frac{2\log_2(1 + M/N)}{Rh} \quad (7)$$

记  $q = |z/z|^2 \frac{2\log_2(1 + M/N)}{Rh}$ , 随着训练的进行,  $q$  应逐渐减小, 当训练到刚开始出现过拟合时,  $q = q$ , 而过拟合参数  $q$  满足  $7 \times 10^{-3} < q < 7 \times 10^{-2}$ , 故此时亦有

$$7 \times 10^{-3} < |z/z|^2 \frac{2\log_2(1 + M/N)}{Rh} < 7 \times 10^{-2} \quad (8)$$

BP 网络用于某一具体问题建模,  $R$ 、 $h$  是确定的; 在训练过程中, 样本集每训练完一遍, 可以计算出  $|z/z|$  和  $M/N$  值. 因此只要在训练过程中, 不断监视  $q = |z/z|^2 \frac{2\log_2(1 + M/N)}{Rh}$  值的变化. 随着训练的进行, 其值不断减小, 若在 BP 网络达到指定精度停止训练前, 总有  $q > 7 \times 10^{-2}$ , 则认为训练过程中, 并未出现过拟合. 若在停止训练前, 已有  $q < 7 \times 10^{-2}$ , 甚至

$q = 7 \times 10^{-3}$ , 则可认为训练过程中已出现过拟合, 尽管训练  
 还未达到指定精度, 也必须停止训练. 可见关系式 (2) 对实际

问题的 BP 网络建模训练过程中防止出现过拟合, 使具有  
 较佳泛化能力有一定指导意义.

表 1 多种不同类型函数的过拟合数值模拟试验的计算结果

函数类型	$n$	$h$	$m$	$m$	$R$	$ z/z $	$ y/y $	$M$	$N$	$q$ ( $1 \times 10^{-2}$ )
线性函数类 $y_1 = a_i x^i$	5	3	117	20	1.0000	0.0310	0.1849	0.4844	0.0106	2.1233
	3	5	60	14	1.0000	0.1513	0.1645	0.5491	0.0131	5.3977
	7	9	78	15	1.0000	0.0265	0.1282	0.5403	0.0130	0.4090
	5	5	88	36	1.0000	0.1152	0.1274	0.4872	0.0117	3.1758
	4	2	80	57	1.0000	0.0441	0.0596	0.4815	0.0148	1.3306
对数函数类 $y_2 = a_i \log(x_i + i)$	5	6	126	15	0.9098	0.0789	0.2462	0.6664	0.0464	2.8073
	3	8	63	45	0.8838	0.0640	0.0702	0.7078	0.0271	0.6049
	4	4	37	35	0.9120	0.1160	0.1883	0.5935	0.0343	5.0218
	7	6	68	17	0.8938	0.1036	0.2197	0.6986	0.0280	3.9874
	4	8	96	48	0.8718	0.0860	0.0590	0.6814	0.0372	0.6218
指数函数类 $y_3 = a_i e^{x_i}$	7	4	105	60	0.9931	0.0652	0.1793	0.5341	0.0241	2.6687
	4	10	58	13	0.9885	0.1039	0.1767	0.4690	0.0200	1.7127
	6	6	60	54	0.9902	0.1028	0.2493	0.4002	0.0207	3.7563
	7	9	123	15	0.9930	0.0459	0.1259	0.5145	0.0208	0.6065
	3	7	100	23	0.9920	0.0362	0.1808	0.4595	0.0117	1.0046
幂函数类 $y_4 = a_i x_i^i$	7	4	118	35	0.9612	0.1413	0.2412	0.3801	0.0373	6.1817
	5	6	75	9	0.9291	0.0898	0.3110	0.4701	0.0227	4.4459
	4	8	56	45	0.9401	0.1980	0.2171	0.3403	0.0337	3.9701
	3	7	30	28	0.9831	0.0780	0.1343	0.3904	0.0207	1.3133
	5	8	82	11	0.9498	0.1051	0.2810	0.4589	0.0229	3.4166
任意函数类	6	10	80	39	0.4510	0.1588	0.2011	0.4310	0.0457	4.7791
	9	8	43	23	0.7364	0.0669	0.1606	0.3790	0.0167	1.6656
	8	13	66	28	0.6033	0.1069	0.2633	0.4688	0.0284	2.9642
	7	11	113	31	0.3722	0.1431	0.2048	0.4269	0.0265	5.8649
	5	9	102	20	0.5248	0.2012	0.2236	0.4710	0.0891	5.0524

### 5 结论

(1) 由于  $|z/z|$  的计算比  $|w|$  的计算简单, 因此本文建立的不确定关系式 (2) 比文献 [13] 建立的关系式 (1) 更直接、更明确、也更简单地揭示了 BP 网络过拟合时的学习能力与泛化能力之间满足的定量关系式.

(2) 关系式 (2) 中的过拟合参数  $q$  的取值范围 ( $7 \times 10^{-3}, 7 \times 10^{-2}$ ) 的区间长度比关系式 (1) 中过拟合参数  $p$  的取值范围 ( $1 \times 10^{-5}, 5 \times 10^{-4}$ ) 的区间长度更缩小.

(3) 虽然关系式 (2) 只是针对 BP 网络提出的, 但关系式形式对于其他任何前向神经网络也应适用, 只是不同的前向网络关系式中某些字母代表的具体含义和确定出的参数  $q$  的取值范围有所不同, 因此, 关系式 (2) 具有普遍意义.

(4) 本文虽已对包括任意复杂函数在内的 5 类函数进行了数值试验, 确定出关系式中参数  $q$  的取值范围. 为进一步检验  $q$  的取值范围. 在此基础上笔者还准备对更多类型函数进行模拟数值试验.

(5) 关系式 (2) 建立过程中, 只考虑了网络结构、函数复杂性和问题规模对学习能力和泛化能力的影响, 并未考虑样本质量、样本代表性、初始权值和先验知识等对泛化能力的影响; 此外, 复相关系数  $R$  并不能完全表征函数 (问题) 的复杂性. 这些问题都有待于进一步深入探索.

### 参考文献:

[ 1 ] Baum E B, Haussler D. What size net gives valid generalization [A]. NIPSI [C]. San Mateo, CA, 1989. 81 - 90.  
 [ 2 ] Moody J E. The effective number of parameters: An analysis of generalization and regularization in nonlinear learning system [A]. NIPS 4 [C]. San Mateo, CA, 1992. 847 - 854.  
 [ 3 ] Barron A R. Approximation and estimation bounds for artificial neural networks [J]. Machine Learning, 1994, (14) : 115 - 133.  
 [ 4 ] Geman S. Neural networks and bias/variance dilemma [J]. Neural Computation, 1992, (4) : 1 - 58.  
 [ 5 ] Cataltepe Z, Abur-mostafa Y. S, Magdorr-Ismael M. No free lunch for early stopping [J]. Neural Computation, 1999, (11) : 995 - 1009.

- [ 6 ] Amari S, Murata N, Muller K R, et al. A symptotic statistical theory of overtraining and cross-validation [J]. IEEE Trans Neural Networks, 1997, 8(5) :985 - 996.
- [ 7 ] 张鸿宾. 训练多层网络的样本数问题 [J]. 自动化学报, 1993, 19(1) :71 - 77.
- [ 8 ] 魏海坤, 徐嗣鑫, 宋文忠. 神经网络的泛化理论和泛化方法 [J]. 自动化学报, 2001, 27(6) :806 - 815.
- [ 9 ] 阎平凡. 人工神经网络的容量、学习与计算复杂性 [J]. 电子学报, 1995, 23(4) :63 - 67.
- [ 10 ] 王晖, 何新贵. BP 网络泛化能力改进研究 [J]. 系统工程与电子技术, 2001, 23(3) :85 - 87, 101.
- [ 11 ] 江学军, 唐焕文. 前馈神经网络泛化性能的系统分析 [J]. 系统工程理论与实践, 2000, 20(8) :38 - 40.
- [ 12 ] 彭汉川, 甘强, 韦钰. 提高前馈神经网络推广能力的若干实际方法 [J]. 计算机工程与应用, 1999, (1) :47 - 48.
- [ 13 ] 李祚泳, 蔡辉. BP 网络过拟合满足的不确定关系式 [J]. 系统工程与电子技术, 2002, 24(9) :94 - 96, 125.

#### 作者简介:



李祚泳 男, 1944 年生于四川省宜宾, 硕士, 教授, 博士生导师, 四川省学术带头人, 主要从事人工神经网络、遗传算法、免疫算法、投影寻踪技术、物元可拓集、粗集理论和蚁群算法等的基础研究和应用基础研究, 已发表论文 250 余篇, 论文被 SCI、EI、CA 检索 50 余次, 被他人引用 200 余次, 获省部级科技进步奖 8 项。

易勇鸷 男, 1979 年生于四川省绵阳, 成都信息工程学院电子系 2002 年本科毕业生。