

BP 网络的过拟合现象满足的测不准关系式*

李祥泳 邓新民

TP18

(成都气象学院大气电子技术研究所, 四川, 成都, 610041)

摘要 揭示了在 BP 网络建模过程中, 出现“过拟合”现象时网络学习能力与推广能力之间满足的一般测不准关系式; 通过数值模拟试验确定出测不准关系式中的“过拟合”参数的取值范围及其与函数类的复杂性的关系; 指出函数类的复杂性可用复相关系数来表示; 给出应用 BP 网络对实际问题的建模过程中, 为了避免出现“过拟合”现象, 提高网络推广能力的“过拟合”判别式。

关键词 BP 网络, 学习能力, 推广能力, 过拟合, 测不准关系, 信息论。

UNCERTAINTY RELATION SUITED TO OVERFITTING OF BP NEURAL NETWORK*

LI Zuo-Yong DENG Xin-Min

(Institute of Atmospheric Electronic Technique, Chengdu Institute of Meteorology, Chengdu, Sichuan 610041, China)

Abstract A general uncertainty relation between learning ability and generalization ability suited to overfitting was revealed in the modeling of BP neural network. Tests of numerical simulation were carried out to determine the field of values of overfitting parameter in the uncertainty relation, and the relations between the value assignment of parameters and the complexity of function category, which can be represented with the aid of complex correlation coefficient. An overfitting discriminant which improves the generalization ability of network was also presented for freeing of overfitting in the modeling of cases using BP neural network.

Key words BP network, learning ability, generalization ability, overfitting, uncertainty relation, information theory.

引言

研究 BP 网络“过拟合”现象出现时的学习能力与推广能力之间的相互关系是 BP 网络建模的基本问题^[1]。对“过拟合”现象的研究, 有人提出了改善推广能力的一些实际方法^[2~6], 但对“过拟合”现象出现时的网络学习能力与推广能力之间满足的解析表达式, 目前在国内外尚无报道。本文将 BP 网络的建模过程与信息的传递过程相类比, 由信息论的一般测不准关系式, 导出了 BP 网络建模出现“过拟合”现象时, 网络的权值变化与对新样本的辨识误差之间满足的测不准关系式。

1 BP 网络学习能力与推广能力间满足的“过拟合”测不准关系式

在信息论中, 根据香农-维纳(Shannon-Wiener)^[7]公式, 一个质量为 m 的物体以光量子形式传递的最大

平均信息量 S 满足一般测不准关系式

$$\Delta P \cdot \Delta X \geq \frac{S h}{2 \log_2(1 + M/N)}, \quad (1)$$

式(1)中, ΔP 和 ΔX 分别为物体的动量变化和位置变化; S 为传递的最大平均信息量; M 为信息平均功率; N 为环境“噪声”的平均功率; M/N 为信噪比, $h = \frac{h}{2\pi}$, h 为普朗克常数。

从信息论观点来看, BP 网络的训练过程是网络从样本获得的信息, 经激活函数作用后, 以权值作载体, 传递并输出信息的过程。在训练过程中, 不断改变网络权值适应从样本获取信息的学习能力。将 BP 网络的建模过程与信息的传递过程相对比, 它们之间的有关量有如下对应关系: 网络对新样本的辨识误差 Δy 相应于观测到的物体的位置变化 ΔX ; 网络的权值变化 ΔW 相应于观测到的物体的动量变化 ΔP ; 从训练样本

* 国家“九·五”重点科技攻关基金(编号 96-911-08-03)资助项目
稿件收到日期 1999-11-16

* The project supported by the National “9·5” Key Foundation of Science and Technology of China(No. 96-911-08-03).
Received 1999-11-16

获取的最大信息量 n 相应于传递的最大平均信息量 S ; 训练样本的网络输出信息的平均归一化值 F 相应于信息平均功率 M ; 训练样本的网络输出信息的方均根误差 E (噪声) 相应于环境“噪声”的平均功率 N ; 输出信号与误差的比值 F/E 相应于信噪比 M/N ; “过拟合”参数 p 相应于普朗克常数 h .

类比信息论中的一般测不准关系式(1), 可以推测: BP网络建模过程出现“过拟合”现象时, 相继训练完毕两遍后, 反映学习能力的网络权值改变量 $|\Delta W|$ 与表征网络的推广能力对新样本辨识的方均根误差 $|\Delta y|$ 之间也满足类似“过拟合”测不准关系式

$$|\Delta W| \cdot |\Delta y| \geq \frac{np}{2\log_2(1+F/E)}, \quad (2)$$

式(2)中, n 为样本因子数, 在样本数 m 和 n 满足 $\log_2 m \geq 2$ 情况下, n 代表信息量. F 和 E 分别为训练完 T 遍后, 训练样本的网络输出信息的平均归一化均值和方均根误差; p 为使测不准关系式(2)成立的待确定的“过拟合”参数. ΔW 为网络出现“过拟合”现象时, 相继训练完毕两遍后网络权值平均改变量. 第 i 个输入节点的权值计算公式为

$$W_i = \sum_{j=1}^H |W_{ij}| \cdot |V_j|, \quad (3)$$

式(3)中, H 为隐节点个数; W_{ij} 为第 i 个输入节点与第 j 个隐节点的连接权值; V_j 为第 j 个隐节点与输出节点的连接权值. 网络相继训练完毕两遍后, n 个输入节点的权值改变的绝对值的和的平均值为

$$\begin{aligned} \Delta W &= \frac{1}{n} \sum_{i=1}^n |\Delta W_i| \\ &= \frac{1}{n} \sum_{i=1}^n |W_i(T) - W_i(T-1)|. \end{aligned} \quad (4)$$

表1 几类不同函数“过拟合”现象出现时的 $|\Delta W|$ 、 $|\Delta y|$ 、 p 和 E

Table 1 $|\Delta W|$, $|\Delta y|$, p and E in the “overfitting” of BP networks of some different function categories

函数类型	$ \Delta W $	$ \Delta y $	p	E
A类	$10^{-5} \sim 10^{-4}$	$10^{-1} \sim 10^{-2}$	$10^{-4} \sim 10^{-5}$	$10^{-1} \sim 10^{-2}$
B类	$10^{-2} \sim 10^{-1}$	$10^{-1} \sim 10^{-2}$	$10^{-1} \sim 10^{-1}$	$10^{-1} \sim 10^{-1}$
C类	$10^{-2} \sim 10^{-2}$	$10^{-1} \sim 10^{-2}$	$10^{-2} \sim 10^{-1}$	$10^{-1} \sim 10^{-2}$

表2 几类不同函数的复相关系数 R^*

Table 2 Complex correlation coefficients R^* of some different function categories

n	A类		B类		C类	
	线性函数	对数函数	幂函数	指数函数	线性逆函数	
3	1.000	0.9839	0.9835	0.8954	0.9051	
3	1.000	0.9786	0.9722	0.8059	0.8378	
5	1.000	0.9780	0.9704	0.7893	0.7889	

2 数值模拟试验

2.1 确定“过拟合”参数值

采用3层BP网络建模进行数值模拟. 选取网络的输出节点 $K=1$ 个, 隐节点数变化范围为: $j=3 \sim 15$; 输入节点数变化范围为 $n=3 \sim 9$. 将模拟的连续函数设计为3大类: A类: 简单函数(线性函数 $y = \sum_{i=1}^n a_i x_i$, 或近线性函数); B类: 较复杂函数(幂函数 $y = \sum_{i=1}^n a_i x_i^i$ 和对数函数 $y = \sum_{i=1}^n a_i \ln x_i$); C类: 复杂函数(指数函数 $y = \sum_{i=1}^n e^{a_i x_i}$, 线性逆函数 $y = 1 / \sum_{i=1}^n a_i x_i$). 对上述每一类函数, 训练样本数 m , 因子数 n , 系数 a_i 和自变量 x_i 在规定范围内随机赋予. 一共进行了数百次任意组合建模试验, 得到上述几类不同复杂性函数BP建模过程中, 当“过拟合”现象出现时, 满足测不准关系式(2)成立的 $|\Delta W|$ 、 $|\Delta y|$ 、 p 和 E 如表1所示.

一般 p 的取值范围为 $10^{-1} \sim 10^{-6}$, 不同类型函数取值范围略有差异. 从式(2)可见BP网络出现“过拟合”现象后, 若继续训练, $|\Delta W|$ 将继续减小, 而 $|\Delta y|$ 必然增大, “过拟合”现象将更加显著.

2.2 函数复杂性的判定

函数的复杂性可以用复相关系数来描述. 通过数值模拟试验得出几类不同函数的复相关系数 R^* 如表2所示. 一般而言, 随着函数复杂性增加, 复相关系数 R^* 逐渐减小, 即 $R_A^* > R_B^* > R_C^*$.

3 “过拟合”判别式

设用BP网络建模的问题中有 m 个训练样本, 每个样本有 n 个因子, 首先求出 n 个因子组合后与 m 个样本集的复相关系数 R^* ; 根据 R^* 大小可以粗略估计

表3 由“过拟合”判别式判断“过拟合”现象出现前后的权值改变的数值检验结果

Table 3 Numerical check results of weight changes around "overfitting" based on "overfitting" discriminant

序号	m	n	J	F_0	F	$ E $	$ \Delta W_1 $	$ \Delta W_2 $
1	70	6	13	$1.0 \cdot 10^{-3}$	$1.90 \cdot 10^{-3}$	$2.23 \cdot 10^{-2}$	$4.11 \cdot 10^{-1}$	$4 \cdot 10^{-4}$
2	20	3	7	$1.0 \cdot 10^{-5}$	$1.62 \cdot 10^{-1}$	$2.61 \cdot 10^{-2}$	$1.66 \cdot 10^{-1}$	$1.7 \cdot 10^{-3}$
3	25	4	4	$1.0 \cdot 10^{-6}$	$1.67 \cdot 10^{-1}$	$3.12 \cdot 10^{-2}$	$2.14 \cdot 10^{-1}$	$2.2 \cdot 10^{-2}$

该问题的复杂性程度,从而确定其函数类型;再根据函数类型选取测不准关系式(2)中的参数 p 的取值范围. 设选取 $p=p_0$, 因为“过拟合”现象刚出现时, 一般有 $|\Delta y| \sim E$. 因此, 在式(2)中, 令 $p=p_0$, $|\Delta y|=|E|$, 则测不准关系式(2)变为

$$|\Delta W| \cdot |E| \geq \frac{nP_0}{2\log_2(1+F/E)} \quad (5)$$

在 BP 网络训练初期阶段, 由于 $|\Delta W|$ 和 $|E|$ 都较大, 不等式(5)总能满足; 随着训练进行, 不等式(5)左边的 $|\Delta W|$ 和 $|E|$ 迅速减小, 而右边变化不大. 当训练进行到某个时候, 上式不再成立, 出现

$$|\Delta W| \cdot |E| \leq \frac{nP_0}{2\log_2(1+F/E)} \quad (6)$$

说明此时已出现“过拟合”现象, 应停止训练. “过拟合”判别式(6)即可作为网络“过拟合”现象出现的判别依据.

用多组不同数据对 BP 网络训练过程中出现从不等式(5)到不等式(6)的转变进行检验. 列出其中 3 组数据, 网络出现“过拟合”现象前后分别满足关系式(5)和(6)的 $|\Delta W_1|$ 和 $|\Delta W_2|$ 值如表 3 所示.

4 结语

本文揭示了 3 层 BP 网络的学习能力与推广能力之间满足的“过拟合”测不准关系式, 确定出“过拟合”参数 p 的取值范围为 $10^{-3} \sim 10^{-6}$; “过拟合”参数 p 的

取值范围随函数复杂性的增加而向 p 值大的方向移动; 而函数的复杂性随复相关系数的减小而增加; 实际问题的 BP 网络建模过程中, 满足“过拟合”判别式时, 应停止训练, 避免出现“过拟合”现象. 多个隐层的 BP 网络, 出现“过拟合”现象时, 原则上应满足类似形式的测不准关系式.

REFERENCES

- [1] YAN Ping-Fan. Some views on the research of multilayer feedforward neural networks, *Acta Electronica Sinica* (阎平凡. 对多层前向神经网络研究的进一步看法, *电子学报*), 1999, 27(5): 82-85
- [2] LIU G Y, YU J B, QIN Y H. A fast BP learning algorithm via a hybrid approach, *J of UEST of China*, 1998, 27(3): 265-268
- [3] Moody J O. The dependence identification neural network construction algorithm, *IEEE Trans. on NN.*, 1996, 7(1): 3-15
- [4] Chen Philip. A rapid supervised learning neural network for function interpolation and approximation, *IEEE Trans. on NN.*, 1996, 7(5): 1220-1230
- [5] Amari S. Natural gradient works efficiently in learning, *Neural Computation*, 1998, 10: 251-276
- [6] HU Yao-Gai, LI Kai-Yang, ZHONG YU-Ning. An improved BP algorithm of neural networks, *J. Wuhan Univ.* (Natural Science Edition) (胡耀垓, 李凯扬, 钟毓宁. 一种改进的神经网络 BP 算法, *武汉大学学报(自然科学版)*), 1999, 45(1): 25-29
- [7] ZHA You-Liang. Information uncertainty principle, *Chinese Science Bulletin*, 1989, 34(1): 86-87