

隐层神经元饱和化与稳健性研究

On the Saturation at the Hidden-Layer Neuron versus the Robustness of the MLFNN

刘光远 邱玉辉

虞厥邦

(西南师范大学计科系 重庆 400715) (电子科技大学光电系)

摘要 本文通过对多层前向网络(MLFNN)中隐层神经元激活函数的强迫饱和化,降低了网络输出对输入的敏感性,进而改善了网络的稳健性(Robustness)。文中给出了相关概念和理论分析,同时以函数逼近问题为例进行了相应的计算机仿真,其结果证明了理论分析的正确性。

关键词 多层前向网络,饱和化,稳健性,敏感性,目标函数

一 引言

尽管多层前向网络使用反向传播算法(BP)训练之后能够解决复杂的模式分类、函数逼近、语音识别等许多问题,但在实际应用中如何进一步提高该类网络的稳健性仍是一个值得研究的问题,一则实际网络参数随环境条件改变要发生变化,二则实际输入信息随时有噪声的加入会影响到网络的实际输出,如何改进前向网络的稳健性已有许多研究报道,比如修剪对输入敏感的权值的方法,衰减大的权值的方法及修正目标函数的方法等等,但将隐层神经元饱和化却是一种新的方法。该方法的基本思想是通过将隐层神经元激活函数(Sigmoid)强制推向饱和,降低输出对于输入的敏感性,达到改善网络稳健性的目的。

二 理论分析及 BP 算法修正

2.1 基本概念

目前,大量的文献把网络的稳健性(或称鲁棒性)与网络的泛化能力混为一谈。为避免混淆,本文首先给出几个基本概念的定义。

(1)稳健性(Robustness) 当输入信息或网络参数发生有限摄动时,网络仍能保持正常或稳定的输入-输出关系特性。

(2)泛化能力 当受训网络接受同一样本集中的非训练样本时网络仍能给出正确的输入-输出关系的能力。

(3)设计点 受训后网络参数的取值点。

(4)饱和度 神经元激活函数在非线性区域的

平缓度,可用该函数的斜率表示。

(5)敏感度 由网络输入变化引起的网络输出变化大小。

前向网络的稳健性依赖于设计点附近系统误差曲面的具体形态。当设计点附近系统误差曲面较为平缓时,则网络的稳健性就好;否则,网络的稳健性就差。而网络的泛化能力依赖于网络的结构及训练样本的特性。

对有 L 层的前向网络,应用链式规则可知其敏感度为

$$\frac{\partial y_k}{\partial x_i} = \sum_{j=1}^{L-1} W_{ij}^{L-1} W_{ji}^{L-2} \cdots W_{ij}^1 f'_{L-1}(y_k) f'_{L-2}(h_{L-1}^j) \cdots f'_1(h_1^j) \quad (1)$$

其中, x_i 和 y_k 分别表示输入向量中第 i 个元素和输出向量中的第 k 个元素, W_{ij}^{l-1} 是第 l 层和第 $l-1$ 层的连接权值, $f'_l(\cdot)$ 是第 l 层非线性函数 $f_l(\cdot)$ 的一阶导函数。存在下列关系:

$$h_l^j = \sum_{i=1}^{x_{l-1}} W_{ij}^{l-1} h_{i-1}^j \quad (2)$$

$$h_l^j = f_l(h_l^j) \quad (3)$$

由(1)式可知,当 $f'_l(h_l^j) \leq 1$ 时, $\partial y_k / \partial x_i$ 很低。这可通过强迫隐层神经元激活函数达到饱和状态来实现。这也就是本文的中心任务。

2.2 理论分析

为简单起见,仅以一个三层前向网络(即只含一个隐层且输出层单元为线性激活函数)为例进行分析,其结果可以说明多隐层的情况。

现设输入 x 的变化为 δx , 由 δx 引起的输出变

化可近似为:

$$\delta y_k = y_k(x + \delta x) - y_k(x) \approx \sum_j \delta x_j \partial y_k / \partial x_j = \sum_j \sum_i W_{ji}^{(l)} f'(\hat{h}_j) W_{ij}^{(l-1)} \delta x_i \quad (4)$$

显然,为了改善网络的稳健性,取较小的 $|\delta y_k / y_k| \approx R$ 是必要的。

$$\text{由 } y_k = \sum_j W_{kj}^{(l)} f(\hat{h}_j) \quad (5)$$

及(4)式得:

$$R = \left| \frac{f'(\hat{h}_j)}{f(\hat{h}_j)} \sum_i W_{ij}^{(l-1)} \delta x_i \right| \quad (6)$$

对模拟量输入和小扰动,有 $|\delta x_i| \ll |x_i|$, 上式变成:

$$R < \frac{f'(\hat{h}_j)}{|f(\hat{h}_j)|} \left| \sum_i W_{ij}^{(l-1)} x_i \right| = f'(\hat{h}_j) | \hat{h}_j / f(\hat{h}_j) | \quad (7)$$

取隐层激活函数为双极性 Sigmoid 函数,即 $f(x) = (1 - e^{-x}) / (1 + e^{-x})$, 则(7)式变为:

$$R \leq R_0 = 2 \hat{h}_j / [\exp(\hat{h}_j) - \exp(-\hat{h}_j)] \quad (8)$$

易知, (8)式 R_0 在 $\hat{h}_j = 0$ 处有极大值,而在 $\hat{h}_j = \pm \infty$ 处有两个极小点。当 $|\hat{h}_j|$ 增大时, R_0 呈指数型减小(如图 1 所示)。因此,可以得出这样的结论:敏感度函数 R 是一个有上界的函数,对于较大的输入变化只能导致更小的网络输出改变。当隐层神经元激活函数被推向饱和区域时,由输入变化引起的输出变化被限制在比零稍大的很小范围内,使网络具有了更好的稳健性。

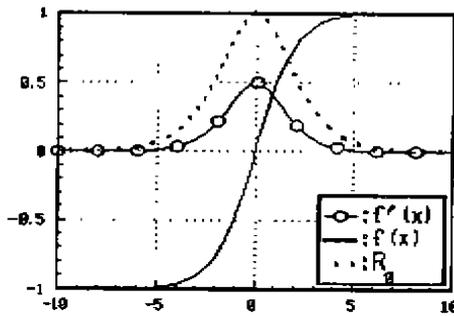


图 1 双极性 f, f' 及 R_0 函数曲线图

2.3 BP 算法的修正

要达到使隐层神经元饱和化的目的,只须在标准 BP 算法的目标函数中增加一个惩罚项,即:

$$E = E_0 + \sum_{i=1}^{l-1} \gamma_i E_i' = \frac{1}{M} \sum_{i=1}^M E_i + \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^{l-1} \gamma_j E_j' \quad (9)$$

其中,标准目标函数 $E_0 = \frac{1}{2N_0} \sum_i (t_i - y_i)^2$ (10)

惩罚项 $E_i' = \frac{1}{N_i} \sum_j f'(\hat{h}_j')$ (11)

这里 M, N_0 和 N_i 分别为训练模式数、输出神经元数和第 l 隐层的神经元数, t_i 和 y_i 分别为第 i 个模式的目标值和输出值,而 γ_i 为第 l 层的惩罚因子,由实际情况确定 ($0 < \gamma_i < 1$), $f'(\hat{h}_j')$ 为激活函数的导数,对双极性 Sigmoid 函数而言,它的值小于等于 0.5 (见图 1)。

在(9)式中加入惩罚项后,极小化标准目标函数的同时也使 E_i' 极小化,将网络推向激活函数的非线性区域,达到了改善网络稳健性的目的。

由于目标函数的修正,原 BP 算法连接权及隐层误差公式应得到相应的修正。

连接权刷新公式为:

$$\Delta W_{ij}^{(l-1)} = -\eta_l \frac{\partial E}{\partial W_{ij}^{(l-1)}} = \eta_l h_{i-1}^{(l-1)} \delta_j^{(l)} \quad (12)$$

式中 $\delta_j^{(l)}$ 为第 l 隐层第 j 个神经元的一般化误差,详细表达式为:

$$\delta_j^{(l)} = -\frac{\partial E}{\partial \hat{h}_j^{(l)}} = \left(\sum_{i=1}^{N_{l+1}} \delta_{i+1}^{(l+1)} W_{ij}^{(l+1)} + \frac{\gamma_l}{MN_{l+1}} \right) f'(\hat{h}_j^{(l)}) \quad (13)$$

对输出层有:

$$\delta_{iL} = \frac{1}{MN_L} (t_{iL} - y_{iL}) f'(\hat{y}_{iL})$$

对双极性函数有 $f'(x) = (1 - f^2(x)) / 2$ 。

三 计算机仿真结果和分析

一般说来,训练之后的网络在应用时往往会有输入噪声改变其输出结果。为了使这种改变尽可能地小,必须增强网络的抗干扰能力,下面的例子就前面所提及的方法说明了这个问题的解决办法。

逼近函数取为

$$y = x / (1 + x^2), x \in [-2, 2] \quad (15)$$

我们在相同的训练参数、网络结构(1—10—1)及训练集(由 50 个数据点)情形下,只改变惩罚因子 γ 进行训练(注意到 $\gamma=0$ 即为标准 BP 算法)。训练终止标准 $E=0.05$, 测试集由与训练集不完全相同的另外 50 个数据构成,网络采用双极性 Sigmoid 函数作为隐层神经元激活函数,输出层取为线性函数 $f(x) = x$, 所加输入噪声为均值由 0 到 100, 方差为 0.1 的高斯噪声,点数为 5 (即 10%), 每种情况下所进行

(下转封四)

(上接第 63 页)

的测试次数是 20, 在 20 次中每次噪声点的加入都是随机的, 因而这种测试方法是可信的, 详细测试结果见表 1。

表 1 对网络输出的测试结果

加噪情况	学习率 $\alpha=0.1$, 动量因子 $\beta=0.7$		
	$\gamma=0.0$	$\gamma=0.05$	$\gamma=0.1$
N(0.0, 0.1)	0.098855	0.027762	0.022098
N(0.0, 1)	0.105323	0.036565	0.028460
N(1, 0.1)	0.097432	0.046507	0.041581
N(5, 0.1)	0.323169	0.288976	0.285157
N(10, 0.1)	0.330256	0.288909	0.285157
N(50, 0.1)	0.337443	0.288909	0.285157
N(100, 0.1)	0.338024	0.288909	0.285157

表中所测之值为网络输出的均方根值, 该值常被用来评价网络的性能。它的定义为:

$$RMS = \frac{1}{M} \sum_{i=1}^M \left(\frac{1}{N_0} \sum_{j=1}^{N_0} (t_i - y_i^j)^2 \right)^{1/2} \quad (16)$$

分析表 1 中的结果可以得出两点: (1) 三种训练情况下随着噪声值的增大其均方根值都有增加, 但在相同噪声幅值输入的情况下, 若 γ 值越大其均方根值越小, 或者说网络越饱和, 其稳健性也越好, 反之亦然。(2) 经过饱和化训练后, 噪声幅度增加到一

定值 ($\gamma=0.05$ 时为 10, $\gamma=0.1$ 时为 5) 后, 网络的输入输出关系便趋稳定, 不再随噪声幅值的增大而变化。这说明所提出的方法的正确性。

结论 研究结果表明, 在原标准 BP 算法中的目标函数项内增加一惩罚项, 可以使隐层神经元激活函数在训练后处于非线性饱和区域, 进而改善了网络的稳健性, 所增加的额外计算量是微不足道的。该方法同时也具有较好的泛化能力, 我们将在今后的工作中对此作进一步的报道。

参考文献

- 1 Jeong D G, Lee S Y. Merging back-propagation and Hebbian learning rules for robust classifications. *Neural networks*, 1996, 9(7)
- 2 Chen D S, Jain R C. A robust back propagation learning algorithm for function approximation. *IEEE Trans on Neural Networks*, 1994, 5(3)
- 3 Manian E V, et al. A neural network approach to non-parametric and robust classification procedure. *IEEE Trans on neural networks*, 1997, 8(2)
- 4 Hertz J, et al. Nonlinear backpropagation: Doing Back-propagation without derivatives of the activation function. *IEEE Trans on Neural Networks*, 1997, 8(6)

计算机科学

(1974年2月创刊)

第26卷第2期(月刊)

1999年2月25日出版

中国标准刊号: ISSN 1002-137X
CN51-1239/TP

定价: 7.50元 国外定价: 5美元

邮发代号: 78-68

发行范围: 国内外公开

主管单位: 国家科学技术部

主办单位: 国家科技部西南信息中心

编辑出版: 《计算机科学》杂志社

重庆市渝中区胜利路132号 邮政编码: 400013

电话: (023) 63500828 传真: (023) 63502473

主 编: 朱宗元

印刷者: 国家科技部西南信息中心印刷厂

总发行处: 重 庆 市 邮 政 局

订购处: 全 国 各 地 邮 政 局

国外总发行: 中国国际图书贸易总公司(北京399信箱)

国外代号: 6210M