

混合激活函数对 BP 算法收敛速度的影响

周 玲 孙 军 袁宇波 丁晓群

(河海大学电气工程学院 南京 210098)

摘 要 研究了不同激活函数选取对 BP 网络收敛速度的影响 ,得出了采用组合激活函数可改善 BP 网络的收敛性的结论 .以电力变压器的人工神经网络故障诊断为例所选取的 TTS(T 表示为双曲正切函数 S 表示为 Sigmoid 函数)的组合激活函数方式 ,具有快速收敛性和较高的故障诊断精度 .
关键词 BP 算法 ;收敛速度 ;激活函数
中图号 TM401.1

反向传播神经网络 ,又称 BP 模型 ,是人工神经网络最常用的模型之一 .BP 模型由输入层、隐含层和输出层组成 ,其中隐含层可以是一层或多层 .各层次的神经元之间互相连接 ,同层次内的神经元之间没有连接 .除输入层神经元节点无激活函数外 ,隐含层和输出层神经元节点均有某种激活函数作为传递函数 ,较好地实现了 BP 网络输入和输出的非线性映射问题 .关于 BP 算法中选取不同激活函数的研究 ,近年来已引起国内外许多学者的关注^[1~3] .

根据生物神经元的特性 ,其神经网络中每个神经元的激活函数不一定相同 .也就是说 ,在同一神经网络中 ,激活函数有多种 .由此可归纳为如下两种情况 (a)在人工神经网络中 ,每个神经元的激活函数不同 (b)在人工神经网络中 ,各层神经元的激活函数不同 .

第 1 种情况比较复杂 ,目前只能从生物角度来加以描述 ,难以用具体的数学模型计算 .所以本文研究了第 2 种情况下不同激活函数对 BP 算法的收敛速度的影响 .

1 激活函数的特性

常使用的激活函数有线性函数、硬限幅函数、双曲正切函数或 S 型函数 .在这里 ,我们选用双曲正切函数和 S 型函数 .它们的公式、图形以及导函数图形如下 :

Sigmoid 函数 $f(x) = (1 + e^{-x})^{-1}$
双曲正切函数 $g(x) = (1 - e^{-2x}) / (1 + e^{-2x})$

从图 1 中可知 :双曲正切函数曲线比 Sigmoid 函数曲线陡峭 ,且在零点附近 ,双曲正切函数的导数值比 Sigmoid 函数的导数值大得多 .也就是说 ,在 BP 网络中采用不同的激活函数时 ,产生的神经元激活值不同 .根据生物神经元的基本特点 ,在隐含层采用双曲正切函数 ,输出层采用 Sigmoid 函数 .这是因为 Sigmoid 从形式上比双曲正切函数更具有“ 柔软性 ” .从生物角度看 ,一个人对远远底于或高于他智力和知识水平的问题 ,往往很难产生强烈的思维反应 ;从数学角度看 ,S 函数具有可微分性 .而且 S 函数更接近于生物神经元的信号输出形式 ,所以选用 S 函数作为 BP 网络的输出函数^[4] .S 函数不但具有可微分性 ,而且具有饱和非线性特性 ,这又增强了网络的非线性映射能力 .若用双曲正切函数作为输出函数 ,则由于其曲线斜率大 ,激励程度比较大 ,会产生过激的响应 ,因而与生物神经网络中的信号输出形式不一致 ,而引起振

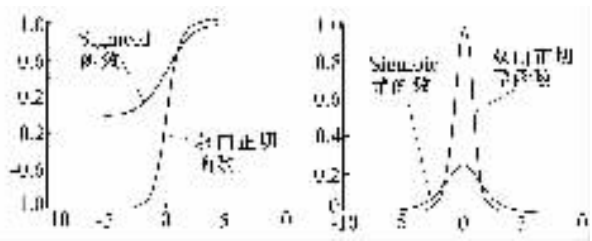


图 1 双曲正切函数和 Sigmoid 函数曲线以及导函数曲线
Fig.1 Curve of hyperbolic tangent Sigmoid and their derivative function

荡或不收敛,并且易陷入局部最小点.

2 算 例

以 BP 网络应用于电力变压器故障诊断的研究为例,验证了采用混合激活函数可改善 BP 网络的收敛性的结论.

具体网络描述:网络结构为 6—2—9,学习率为 0.2,网络的总误差目标值为 0.005,初始权、阈值是随机的.分别改变每层的激活函数,每种结构训练 5 次.SSS 表示 3 层全用 Sigmoid 函数,SST 表示第 1 层、第 2 层采用 Sigmoid 函数,第 3 层采用双曲正切函数,等等,总共有 8 种激活函数混合的网络结构.训练所得结果见表 1.

表 1 网络中不同激活函数时收敛情况
Table 1 The speed of convergence results using different activation function in network

训练次数	SSS	TTS	STT	SST	TSS	TST	STS	TTT
1	8 949	1 385	振荡	振荡	2 730	振荡	3 889	振荡
2	8 367	1 513	振荡	振荡	2 450	振荡	2 939	振荡
3	9 392	1 959	振荡	振荡	3 243	振荡	2 983	振荡
4	8 072	1 789	振荡	振荡	2 609	振荡	1 851	振荡
5	9 636	1 634	振荡	振荡	3 375	振荡	1 895	振荡
平均	8 883	1 656	振荡	振荡	2 881	振荡	2 711	振荡

分析表 1 可知:(a)若网络只采用单一的激活函数,用 S 函数时网络收敛到目标值的迭代次数为 7 000 次以上,用双曲正切函数时,网络不能收敛到目标值.(b)若网络混合使用两种激活函数,则收敛速度出现两极分化,要么迭代次数明显减少,要么出现不收敛.其中 S 为输出函数时迭代次数在 4 000 次以内,但当 T 为输出函数时,基本呈振荡型,出现这种情况是由于网络收敛到局部极小点所致.(c)按平均收敛次数增大的顺序为 TTS,STS,TSS,SSS,即以 TTS 型网络的学习收敛速度最快.另外,由理论分析和仿真计算还得出以下几点普遍性的规律:(a)双曲正切函数能够加快迭代运算的收敛速度.(b)如果要避免 T 函数带来收敛性问题,可降低学习率(比如取学习率为 0.001),但会导致网络的学习收敛速度非常慢,将失去选用混合激活函数的意义.(c)网络输入层采用何种激活函数与训练问题目标值的选取有很大的关系,目标值取值范围应在所取激活函数的值域范围之内.例如:目标值取值范围为 0~1 之间,输出函数选 S 函数较佳;目标值取值范围为 -1~1(或 -0.5~0.5)之间,输出函数选双曲正切函数较佳.

3 结 论

在 BP 网络中层与层之间的激活函数,将直接影响 BP 网络的收敛速度,因此,针对具体问题,网络每层可以选择最合适的激活函数(不只局限于双曲正切函数和 Sigmoid 函数),通过训练来确定最佳的结构.

参 考 文 献

1 向国权,董道珍.BP 模型中激励函数和改进的网络训练法.计算机研究与发展,1997(2):113~117
2 潘丹,罗干英,黄茜等.神经网络活化函数的研究及其应用.华南理工大学学报,1998(4):49~53
3 薛家强,黄石生.BP 神经网络优化训练技术的研究.华南理工大学学报,1998(4):21~23
4 王伟.人工神经网络原理.北京:北京航空航天大学出版社,1995.115~138

Effects of Combined Activation Function on BP Algorithm's Convergence Speed

Zhou Lin Sun Jun Yuan Yubo Ding Xiaoqun
(College of Electrical Engineering, Hohai Univ., Nanjing 210098)

Abstract This paper studies the influence of different activation function on the speed of convergence of BP algorithm, and reaches a conclusion: The combined activation function can improve the speed of convergence of BP algorithm. An example of an artificial neural network(ANN) approach to detecting the faults of transformers is given. The combination of TTS activation functions has quick convergence and high precision in fault diagnosis.

Key words BP algorithm; speed of convergence; activation function