

• 工程应用 •

19 110-115

BP模型的改进算法及其 在负荷预测中的应用

TM715

An Improved Learning Algorithm for BP Model
and its Application to Load Forecasting in Power Systems

罗春雷

Luo Chunlei

孙洪波

Sun Hongbo

徐国禹

Xu Guoyu

(重庆大学电气工程系, 重庆, 630044)

摘 要 分析了BP模型学习算法——累积误差逆传播算法在接近极小点时收敛速度变得异常缓慢的原因,并通过对连接权值的调整量引入权重系数,提出了一种改进的BP模型学习算法,大大加快了收敛速度,提高了收敛性。还利用提出的改进算法对某省中期负荷进行了预测,算例结果表明了该算法的有效性。

关键词 电力系统; 负荷预测 / BP模型; 累积误差逆传播算法
中国图书资料分类法分类号 TM714

学习算法

ABSTRACT This paper analyses the cause of very slow learning rate during the time of approaching the minimum point in conventional BP algorithm, by introducing weight factor into the modification of the connection weight, an improved learning algorithm for BP model is presented, it can significantly increase the learning rate and improve convergent behaviour. As an illustrative application example, the numerical result of the medium-term load forecasting is given.

KEYWORDS power system; load forecasting / BP model; accumulated error back propagation algorithm

0 引 言

1985年, Rumelhart 和 McClelland 提出了误差逆传播模型, 或称逆传播 (Back Propagation) 模型, 简称 BP 模型。BP 模型实现了多层人工神经网络的设想, 其基本网络是三层前向网络, 包含一个输入层, 一个输出层和中间的隐蔽层。假定 BP 网络中隐蔽层的单元数可以根据需要自由设定, 那么这样一个三层网络便可以实现以任意精度近似任何连续函数^[1]。这使得众多现实世界中的问题可以得到解决, 如模式识别、预测、控制、函数拟合等。但是, BP 模型本身尚有许多不足之处, 主要有^[2]:

1) 由于它本质上是一种非线性优化问题,这就不可避免地会遇到最常见的局部极小。

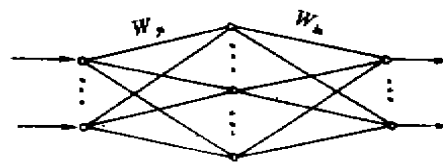
2) BP 学习算法的学习因子常凭经验给定,这使得加速收敛与抑制振荡之间的矛盾不能得到有效解决。

为了解决这些问题,不少人作了努力,其中较为成功的是:Rumelhart 等人用给梯度加上惯性项(反映前一次连接权的变化量)来滤除代阶函数曲面上的高频分量,避免了学习过程的振荡。总之,在误差传播学习过程中加速收敛是最花费精力的内容之一,同时也是值得研究的重要内容。

笔者针对 BP 模型学习算法中,每一连接权值在各个样本中的变化量对整个样本集误差的影响各不相同,分别给予不同的权重系数,从而加速收敛。同时与变学习因子相结合,进一步大大加快学习速度。笔者将该改进算法应用于某省中期负荷预测,并与 BP 模型标准学习算法进行了比较,算例结果验证了所提方法的有效性。

1 BP 模型及其学习算法

BP 模型^[1~3]是一种多层前馈网络,如图中所示,除输入节点层和输出节点层以外,还有一个或多个隐蔽层。相邻两层层间都有连接,且每个连接都对应于一个权值,隐蔽层和输出层的每个神经元都有一个阈值。



BP 模型网络结构图

BP 模型学习可分为两步。在第一步中,对于给定的网络输入,通过现有连接权值将其正向传播,经各个神经元节点的传递函数作用后,获得其实际输出(激活值),节点函数通常使用 S 型(Sigmoid)函数,其表达式为:

$$F(x) = 1/(1 + e^{-x}) \quad (1)$$

在第二步中,先计算出输出层各神经元的一般化误差,这些误差逐层向输入层方向传播,以便获得各层神经元的参考误差,从而调整相应的连接权和阈值,权值及阈值的调整可由以下各式确定:

$$W_{pj}(t+1) = W_{pj}(t) + \Delta W_{pj}(t+1) + \alpha \Delta W_{pj}(t) \quad (2)$$

$$\theta_j(t+1) = \theta_j(t) + \Delta \theta_j(t+1) + \alpha \Delta \theta_j(t) \quad (3)$$

式中, $W_{pj}(t)$ 和 $\theta_j(t)$ 是训练过程中第 t 次循环结束后得到的,从节点 i 到节点 j 之间的权值和节点 j 的阈值; $\Delta W_{pj}(t)$ 和 $\Delta \theta_j(t)$ 是第 t 次循环结束后得到的对应的调整量; α 称为冲量因子,其对应的“惯性项”有抑制振荡的作用。

基于梯度下降法可推导出权值和阈值的调整量为:

$$\Delta W_{pj} = \eta \delta_j O_i \quad (4)$$

$$\Delta \theta_j = \eta \delta_j \quad (5)$$

当节点 j 为输出层的节点时:

$$\delta_j = O_j(1 - O_j)(t_j - O_j) \quad (6)$$

当节点 j 为隐蔽层的节点时:

$$\delta_j = O_j(1 - O_j) \sum_k (\delta_k W_{kj}) \quad (7)$$

式中, η 为学习因子; O_i 为节点 i 的输出值; k 取 j 节点所在层之上一层的所有节点。(4) 式和 (5) 式称为推广的 δ 规则。

重复上述过程, 直到样本集偏差达到允许值。这里样本集偏差常指最小二乘平方和, 即代价函数:

$$E = \frac{1}{2} \sum_p \sum_j (t_{pj} - o_{pj})^2 = \sum_p E_p \quad (8)$$

式中, p 表示样本; j 表示输出层节点; t_{pj} 为节点 j 第 p 个样本的期望值; o_{pj} 为对应的实际计算输出值; E_p 为样本 p 的识别误差。

BP 模型的学习算法如下:

步骤 1 将所有权值和阈值赋初值, 并给出训练样本集:

步骤 2 由输入、权值和阈值计算输出值, 当样本集的偏差小于给定允许值时训练结束, 否则转步骤 3;

步骤 3 根据输出计算值与期望值之间的偏差, 从输出层反向逐层调整权值与阈值, 直至输入层, 调整公式为 (2) ~ (7) 式, 转回步骤 2。

2 改进的 BP 模型学习算法

从 (6) 式可以看出, 当输出的计算值与期望值很接近时, δ_j 变得很小, 这就使得 ΔW_{jk} 在各样本中相差不大, 则权值不能朝着产生误差的主要方向快速修正, 收敛速度因此很慢。这正是由于 BP 模型学习算法实质上使用的是梯度下降法, 而梯度下降法在接近极小点时不可避免地会遇到收敛速度很慢等问题。

通常, BP 学习算法是指累积误差逆传播学习算法, 这样的学习算法使整个样本集的误差逐渐减小, 连接权值和阈值的变化正比于全部样本集上对应的调整量之和, 即权值和阈值的调整量为:

$$\Delta W_{jk} = \sum_p \Delta p W_{jk} \quad (9)$$

$$\Delta \theta_j = \sum_p \Delta p \theta_j \quad (10)$$

由于样本集的任意性, 使得在 BP 网络的整个训练过程中, 连接权值在各样本中的调整量 $\Delta p W_{jk}$ 各不相同, 而且 $\Delta p W_{jk}$ 对该样本识别误差 E_p 的影响, 也随样本的不同而各不相同, 即对整个样本集误差的影响各不相同。如果我们给影响样本集误差较大的调整量以较大的权重系数, 给影响较小的以较小的权重系数, 使调整过程朝着产生误差的主要方向进行, 从而很快减小样本集误差、加速收敛, 影响程度的大小可由代价函数对连接权值调整量的偏导数值来表示, 即由代价函数对连接权的二阶导数值来表示。由于篇幅有限, 下面仅给出推导

结果。

当连接权值 W_{ji} 为隐蔽层节点 i 至输出层节点 j 的连接权值时:

$$\frac{\partial^2 E_p}{\partial W_{ji}^2} = o_{ji}(1 - o_{ji})[o_{ji}(1 - o_{ji}) + (t_{ji} - o_{ji})o_{ji} - (t_{ji} - o_{ji})(1 - o_{ji})]o_{ji}^2 \quad (11)$$

当连接权值 W_{ji} 为输入层节点 i 至隐蔽层节点 j 的连接权值时:

$$\begin{aligned} \frac{\partial^2 E_p}{\partial W_{ji}^2} = & o_{ji}(1 - o_{ji})\{(2o_{ji} - 1) \sum_k (\delta_{jk} w_{ki}) + o_{ji}(1 - o_{ji}) \sum_k o_{jk}(1 - o_{jk})w_{ki}^2 \\ & \times [o_{jk}(1 - o_{jk}) - (t_{jk} - o_{jk})(1 - o_{jk}) + (t_{jk} - o_{jk})o_{jk}]\}o_{ji}^2 \end{aligned} \quad (12)$$

上式中, k 为输出层节点。

由于一个三层 BP 网络已能映射任何非线性函数,故我们没有给出多层 BP 网络的推导结果,实际上根据同样道理仍可推导出。

为叙述方便,令:

$$\frac{\partial^2 E_p}{\partial W_{ji}^2} = h_j \quad (13)$$

h_j 随样本不同而不同,则可以得到一个序列 $\{h_j\}$, h_j 反映了 $\Delta p W_{ji}$ 对整个样本集误差影响的相对大小。取所对应的权重系数为 ω_j , 其中:

$$\max\{h_j\} = h_M \rightarrow \omega_M \quad (14)$$

$$\min\{h_j\} = h_m \rightarrow \omega_m \quad (15)$$

ω_M 及 ω_m 分别为最大最小权重系数,而介于 h_M 和 h_m 之间的 h_j 所对应的 ω_j , 则依 h_j 的相对大小按比例选取 ω_M 和 ω_m 之间的值。在求出 ω_j 后,即可对 $\Delta p W_{ji}$ 加权,从而得到加权后的调整量:

$$\Delta W_{ji} = \sum_j \Delta p W_{ji} \times \omega_j \quad (16)$$

从(11)、(12)、(14)、(15)式可知,权重系数的选取,可直接由(7)、(8)两式中各项的值计算出,故改进 BP 模型学习算法每次循环的计算量与标准 BP 算法相比,并不会增加太多。而在要求样本集误差很小(这正是通常所希望的)时,改进算法需较少的训练时间,具有明显的优越性。

对于改进的 BP 算法,作以下几点说明:

1) 对调整量进行加权,对整个 BP 网络训练过程有加速的作用,特别在接近极小点时表现得更为理想。

2) 因初始连接权值、阈值以及样本集是任意给定的,学习因子及冲量因子均凭经验选

取,故 BP 网络在学习刚开始时可能出现振荡,如果此时引入权重系数,将会使振荡加剧。同时梯度法在开始时下降低效果较为明显,故可设置一误差判别参考值 ξ (如 0.15),仅在样本集误差小于该参考值时,才开始将连接权值在各样本中的调整量乘上不同的权重系数。

3) 为了在乘上权重系数后不致于破坏梯度法的正常使用,权重系数的初值不能过大或过小,可取值在 1.0 左右。在刚开始加权时令 ω_m 为 1.0, ω_w 比 ω_m 略大(如 1.2),当样本集误差减小至 0.1ξ 后,可将 ω_w 加大,而 ω_m 不变(如 ω_m 仍为 1.0,而让 ω_w 增大为 1.5),同时令 ξ 为 0.1ξ 。

4) 从式(4)可看出,学习因子 η 对连接权值的调整量亦有影响,即对收敛速度有影响。然而精确地选取 η 是十分困难的。通常,标准 BP 算法取一较小值。然而在对连接权值的调整量引入权重系数后,则可适当增大 η 的值,从而进一步加速收敛。如在增大 ω_w 时,使 η 变为 1.1η 。

5) 对于学习过程中的阈值也可同样进行加权,只需将阈值 θ_i 看作与某个输出值恒定为 1 的节点相连的“权值”, θ_i 则可以与其它连接权值一样进行上述修正。

3 应 用

电力负荷预测是制定系统规划和运行决策的前提与基础。它的实质是利用以往的数据资料找出负荷的变化规律,从而预测出电力负荷在未来时期的变化趋势及状态。人们已提出了许多预测方法,其中,回归分析法是发展最为成熟、应用最为广泛的一种预测方法^[4],但回归本身存在选择回归线类型较困难等问题。为此,笔者应用 BP 网络来映射预测变量与相关变量之间的关系,从而进行电力负荷预测。

利用所提改进算法,对文献[4]所给出的某省电力负荷进行了预测。影响用电量的因素很多,这里主要选用农业总产值、轻工业总产值及重工业总产值作为相关因素。限于篇幅表 1 仅给出了该省 1981 年至 1984 年各年用电量以及相关变量之值。以该省 1966 年至 1980 年的数据作为历史样本,预测 1981 年至 1984 年的用电量。用已知的相关变量作输入量,用电量作输出量训练 BP 网络。利用本文所提改进算法进行预测,预测结果见表 2。在训练 BP 网络时,将表 1 中各值除以一可能最大值(如 500),以满足 BP 网络的输出值应在区间 $[0,1]$ 内,而输入值不宜过大或过小,从而避开 S 型曲线饱和特征。为了说明权重系数以及学习因子对改进学习算法收敛性的影响,这里采用相同训练参数,用四种方法来训练样本。方法一为标准 BP 算法;方法二为仅增大学习因子;方法三为仅引入权重系数,方法四为本文的改进算法。四种算法的比较结果见表 3。

表 1 各年用电量及工农业总产值

年 份	农业总产值 (亿元)	轻工业总产值 (亿元)	重工业总产值 (亿元)	用电量 (亿 kW·h)
1981	208.21	284.54	181.05	173.0294
1982	234.02	298.13	204.78	184.4195
1983	254.02	298.13	204.78	184.4195
1984	299.46	393.27	286.81	215.5823

表 2 预测结果

年 份	预测值 (亿 kW·h)	绝对误差	相对误差 (%)
1981	173.0715	-0.042	-0.24
1982	185.3885	-0.969	-0.525
1983	200.5814	+1.685	-0.847
1984	220.2368	-4.654	-2.159

表 3 四种方法训练结果比较

训练精度	学 习 次 数			
	方法一	方法二	方法三	方法四
0.000100	5424	3908	1575	1281
0.000087	10000	7559	1576	1282
0.000085	(未收敛)	10000	1576	1282
0.000020	(未收敛)	(未收敛)	1586	1289

由表 2 可见,预测最大相对误差为-2.159%,较之文献[4]中回归分析法预测结果准确,说明用人工神经网络进行电力负荷预测是可行的,有效的。

由表 3 可以看出,对连接权值调整量引入权重系数,能大大加快收敛速度,同时与变学习因子相结合能进一步加快收敛。从表中还可看出,对标准 BP 算法存在的不收敛问题,用本文的改进算法亦能得到改善。

4 结 论

1) BP 模型学习算法在期望值与实际计算值很接近时,权值及阈值的调整量不能充分反映出产生误差的主要方向,大大降低了学习速度,以至于近乎停顿状态。

2) 通过对各样本的调整量用不同的权重系数进行加权,反映出产生误差的主要方向,并改变学习因子,大大加快了收敛速度,同时亦使不收敛现象得到了改善。

3) 对改进的 BP 模型算法,在训练过程中适时地开始加权,以及适当地选取最大权重系数和最小权重系数、增大学习因子,都将进一步减少训练时间。

参 考 文 献

- 1 靳蕃. 神经网络与神经计算机. 成都:西南交通大学出版社,1990. 40~45
- 2 中国科学技术大学. 神经网络及其应用. 合肥:中国科学技术大学出版社,1991. 87~91
- 3 包约翰. 自适应模式识别和神经网络. 北京:科学出版社,1991. 54~58
- 4 刘晨晖. 电力系统负荷预报理论与方法. 哈尔滨:哈尔滨工业大学出版社,1988. 106~108