前馈神经网隐层节点的动态删除法*

陆系群 余英林

(华南理工大学无线电与自动控制研究所·广州,510641)

摘要: 本文首先针对 BP 算法中存在的缺陷对误差函数作了简单的修改,使网络的收敛速度比原来的大大提高. 此外本文提出了一种基于线性回归分析算法来确定隐层节点数. 当已训练好的网络具有过多的隐层单元,可以用这种算法来计算隐层节点输出之间的线性相关性,并估计多余隐层单元数目,然后删除这部分多余的节点,就能获得一个合适的网络结构. 计算机模拟实验结果表明,用这种方法来删除隐层中多余的节点是有效的.

关键词: BP 算法; 误差函数; 线性相关; 删除隐层节点数

1 引 言

BP 算法作为多层前馈神经网络的学习算法已应用到许多方面,如模式识别,函数逼近,数据压缩等等. 但是 BP 算法并不是完美无缺,如在学习过程中由于最速梯度下降法本身的缺陷,网络经常陷入一个局部最小点,学习速度慢等等. 为了克服这些缺点,有两条途径,一是改进 BP 算法,二是针对不同的问题选取一个合适的网络结构.

本文在第二部分针对 BP 算法中存在问题,对误差函数作了简单的修改,使收敛速度得到大大提高. 针对网络的结构,本文提出了一种基于线性回归分析算法来动态地删除多余的隐层点. 当已训练好的网络具有过多的隐层单元时,我们根据隐层单元输出之间的线性相关性删除那些多余的节点,就能得到一个合适的网络结构. 本文第四部分给出了计算机模拟实验结果. 最后部分是结论.

2 修正 BP 算法

BP 算法的一个较突出的缺点是学习速度慢,原因是多方面的,如与网络的结构有关,与学习算法本身存在的缺点有关,我们知道 BP 算法中的误差函数是

$$E = \frac{1}{2}\sum_{p=1}^{N_p}\sum_{k=1}^{N_r}(Y_{k,p} - T_{k,p})^2. \tag{1}$$

其中 p 表示输入样本序号, N_p 为样本总数; Y_{kp} 是在第 p 个样本输入时输出层单元 k 的实际输出, T_{kp} 是它的期望输出, N_r 是输出层单元总数. 在学习过程中逐步调整权重,降低 E,权重的调整量与 E 对权重的偏微分成正比. 权重的调整量都包含有因式 $(Y_k - T_k)Y_k(1-Y_k)$,它是输出单元 k 的误差信号,由输出层反向传播到隐含层.

我们发现当输出层单元 k 的实际输出 Y_k 接近于 0 或 1 时,误差信号中的因子式 $Y_k(1-Y_k)$ 使得误差信号变得很小,这时如果输出层单元 k 的实际输入 Y_k 与期望输出值 T_k 相差很大时,却没有产生强的误差来修正权重,从而延长了学习过程,另外由于激励函数 $f(x) = \frac{1}{1+e^{-x}}$ 是一个饱和函数,当它趋于饱和状态时,导数就接近于零,从而造成收敛速度减慢.

---

* 攀登计划项目,国家自然科学基金资助项目.

本文于 1995 年 5 月 6 日收到. 1996 年 3 月 6 日收到修改稿.

由此我们觉得应将 $Y_k(1-Y_k)$ 因子从误差函数对权重偏微分的结果中去除，于是把 BP 算法的误差函数修正为

$$E=-\sum_{p=1}^{N_p}\sum_{k=1}^{N_p}[T_{pk}\ln Y_{pk}+(1-T_{pk})\ln(1-Y_{pk})]. \tag{2}$$

E 对 Y_k 的偏微分是

$$\frac{\partial E}{\partial Y_k}=\frac{1-T_k}{1-Y_k}-\frac{T_k}{Y_k}=\frac{Y_k-T_k}{Y_k(1-Y_k)}. \tag{3}$$

这样权重的调整量中就消去了 $Y_k(1-Y_k)$，现在反传回去的误差信号与输出节点的期望输出与实际输出之差成正比. 这就是说从一开始学习就快速接近期望输出量并在收敛阶段也接近期望输出，从而提高了学习速度和收敛速度，在本文第四部分计算机模拟实验结果表明，情况正如所预料的那样.

3 用线性回归分析法动态删除多余隐含层节点

文献[1]已证明了具有单隐层的神经网络能够逼近任意非线性函数. 而建立在单隐含层网络上的方法可以推广到多层网络上，因此本文只对单一隐含层网络作研究.

我们知道当隐层节点个数过多会造成网络结构庞大，网络推广能力降低；如果隐层节点数过少，则不能对所要解决的问题形成一个好的模型. 所以选择一个合适的网络结构是网络优化训练的一个重要因素.

文献[1]中已证明了当一个网络具有过多隐层节点时，一些隐层节点的输出之间存在线性相关性. 文献[2]提出了一种根据隐层节点相关系数及样本分散度来合并同一隐层中相关系数过高的两个隐层节点和删除不起作用的隐层节点. 但是一个隐层节点和另一个隐层节点具有线性相关性并不能说明它与其它隐层节点也有线性相关性，这时若把该节点删除掉可能会破坏原来已形成的映射关系. 基于以上这两点. 本文提出了一种用线性回归分析来找出隐层节点输出之间的线性相关部分，然后删除与之相应的隐层单元，最后得到一个合适的网络结构.

3.1 算法原理

设已训练好的网络有 M 个隐层节点，我们先拿第一个隐层节点作为例子，若隐层的第一个节点的输出与后面 $M-1$ 个节点的输出有线性关系，那么第一个节点的输出 Y_1 可以用后面 $M-1$ 个节点的输出的线性组合来逼近：

$$\widetilde{Y}_{1,M-1}=a_{10}^{(1)}+a_{12}^{(1)}Y_2+\cdots+a_{1M}^{(1)}Y_M. \tag{4}$$

若第一个节点输出与后面 $M-2$ 个节点的输出有线性关系，那么第一个节点的输出 Y_1 可以用后面 $M-2$ 个节点的输出的线性组合来逼近：

$$\widetilde{Y}_{1,M-2}=a_{11}^{(2)}+a_{13}^{(2)}Y_3+\cdots+a_{1M}^{(2)}Y_M. \tag{5}$$

以此类推直至最后一个隐层节点的输出若与第一个隐层节点的输出有线性关系，则可表示为

$$\widetilde{Y}_{11}=a_{1,M-2}^{(M-1)}+a_{1M}^{(M-1)}Y_M. \tag{6}$$

其中 $\widetilde{Y}_{1j}$ 是第1个隐层节点输出的估计值，$(j=M-1,M-2,\cdots,1)$，$Y_i(i=1,2,\cdots,M)$ 是第 i 个隐层单元的实际输出，系数 a 是用线性回归分析法来确定的回归系数.

如果第1个隐层单元的输出与后面其他节点的输出有很强的线性相关性，即 $\widetilde{Y}_{1j}(j=M-1,M-2,\cdots,1)$ 全都在高精度上逼近 Y_1. 为了衡量 $\widetilde{Y}_{1j}$ 和 Y_1 之间的相似性，用相关系数 C_{1j} 来表示，其定义如下：

$$C_{1j}=M_{1j}^2/(S_{1j}\widetilde{S}_{1j}), \tag{7}$$

$$M_{1j}=\sum_{i=1}^{N}(Y_{1i}\widetilde{Y}_{1j})-(\sum_{i=1}^{N}Y_{1i}\sum_{i=1}^{N}\widetilde{Y}_{1j})/N, \tag{8}$$

$$S_{1j} = \sum_{i=1}^{N} Y_{1i}^2 - (\sum_{i=1}^{N} Y_{1i})^2/N, \tag{9}$$

$$\widetilde{S}_{1j} = \sum_{i=1}^{N} \widetilde{Y}_{1i}^2 - (\sum_{i=1}^{N} \widetilde{Y}_{1i})^2/N. \tag{10}$$

其中 N 为网络训练样本的总数，$Y_{1i}(i=1,2,\cdots,N)$ 为第 i 个训练样本输入时第 1 个隐层节点的输出；由于有 N 个训练样本，所以第 1 个隐层节点的输出的估计值 $\widetilde{Y}_{1j}$ 也有 N 个，$(j=M-1,M-2,\cdots,1)$.

C_{1j} 的值是在 0 与 1 之间，若 $C_{1j}=1$ 时，Y_1 完全可以用其它隐层节点的输出的线性组合来表示. 若 $C_{1j}=0$ 时，那么 Y_1 与其它隐层单元的输出无线性关系. 但在实际运算中 C_{1j} 不可能完全为 0，我们只能以 C_{1j} 小于某个小数 ϵ 为标准. 本文算法的基本思想是：若 $C_{1j}(j=M-1,M-2,\cdots,1)$ 都大于 ϵ，那么把第一个隐层节点删除；若 C_{1j} 有一个小于 ϵ 值，那么保留第一个隐层节点，然后再从第二个节点开始，重复上述运算过程，直至最后一个节点.

其算法步骤如下：

1）输入训练样本，并给隐层赋予过多的单元数 M.

2）用修正 BP 算法来训练网络，直至收敛.

3）再把训练样本输入到已收敛的网络中；计算隐层节点的输出 $Y_1,Y_2,\cdots,Y_M$.

4）$k \leftarrow 1$.

5）是否已分析到最后一个隐层节点，若是则走第 12 步；若不是则走下一步.

6）用 $Y_k = a_{k,M-j-1}^{(M-j)} + a_{k,M-j+1}^{(M-j)} Y_{M-j+1} \cdots + a_{k,M}^{(M-j)} Y_M,(j=M-1,M-2,\cdots,1)$，作线性回归分析求出系数 a.

7）用(4)、(5)、(6)等式来计算 $Y_k Y_k$ 的估计值 $\widetilde{Y}_{kj}$.

8）计算 $C_{kj},(j=M-1,M-2,\cdots,1)$.

9）若 C_{kj} 都大于 ϵ，则走第 10 步；若 C_{kj} 中有一个小于 ϵ，则走第 11 步.

10）删除第 K 个节点，$M \leftarrow M-1$.

11）$k \leftarrow k+1$，回到第 5 步.

12）结束.

4　计算机模拟实验

4.1　修正 BP 算法性能的测试

在这一节中，我们用学习异或问题做例子，从网络的学习次数及训练误差方面对修正 BP 算法与原来的 BP 算法作比较. 在这个训练过程中，隐层节点数是固定不变的. 隐层单元数为 2 个，误差精度为 0.001，若学习次数超过 1000 次就认为网络学习失败. 在权重初值不同的情况下重复运行算法各 10 次，修正后的 BP 算法都能收到所要求的误差精度，而原来的 BP 算法在 10 次中有一次学习次数已达 1000 次而未能收敛到所要求的精度. 实验结果如表 1 所示.

表 1　XOR 问题模拟结果

算　法	平均学习次数	训练误差
修正 BP 算法	115	0.000927
原有 BP 算法	527	0.000999

表 2　本文方法示例

	训练次数	残留误差
标准 3—3—1	280	0.000085
3—7—1→3—6—1	354	0.000081
3—6—1→3—5—1	381	0.000090
3—5—1→3—4—1	160	0.000085
3—4—1→3—3—1	197	0.000089

4.2 用线性回归分析法动态删除多余隐层节点

我们用一个三位奇偶校验来测试线性回归分析动态删除隐层节点算法的性能. 初始网络有 7 个隐层节点,逐个往下减,直至隐层节点数为 3 个,得到最佳网络结构. 训练结果如表 2 所示.

为了对照起见,表中第一栏列出了只有 3 个隐层节点固定结构时训练次数和残留误差. 从表中可以看出总共训练次数为 1000 次左右就能得到一个合适的网络结构.

5 结 论

本文针对 BP 算法中误差函数的内在缺陷,对误差函数作了简单的修正. 并提出了用线性回归分析来完成动态删除隐层多余节点. 计算机模拟实验结果表明,用这种方法来删除隐层中多余的节点是有效的.

参 考 文 献

1 Funahashi, K.. On the approximate realization of continuous mappings by neural networks. Neural Networks, 1989, 2: 183—192

2 雷鸣,朱心飚,尹申明,杨叔子. 自构形神经网络及其应用. 计算机科学, 1994, 21(1): 52—74

3 David Hrycei. A new algorithm for hidden and input layer pruning. Neural Network World, 1994, 1: 19—35

A Method of Dynamic Pruning the Hidden Layer Nodes in A Feedforward Neural Network

LU Xiqun and YU Yinglin

(Insititute of Radio & Automation, South China University of Technology · Guangzhou, 510641, PRC)

Abstract: In this paper, a modification to the error function of the BP algorithm is proposed, and the convergent rate of the networks is improved. After that an algorithm based on linear regression analysis determining the number of hidden nodes is proposed. By means of a trained network with too many hidden nodes, we calculate the amount of linear correlation between the hidden nodes' outputs and estimate the redundant number of hidden nodes, then pruning the redundant hidden nodes from the initial one, then an appropriate structure can be obtained. The computer simulation has shown that the method is valid in pruning the redundant nodes in the hidden layer.

Key word: BP algorithm; error function; linear correlation; pruning; number of hidden nodes

本文作者简介

陆系群 1970 年生. 1994 年毕业于杭州大学电子系,获硕士学位. 现为华南理工大学无线电系博士研究生. 主要研究方向为信号处理与神经网络.

余英林 1932 年生. 1953 年毕业于华南工学院. 1961 年 12 月在中国科学院电子所副博士研究生毕业. 现在华南理工大学无线电系工作,博士生导师. 从事图象处理、模式识别、神经网络方面研究,在国内外发表论文数十篇.