

确定前向神经网络隐层节点数的模糊聚类分析法

许世刚 索丽生

(河海大学水利水电工程学院 ,江苏 南京 210098)

摘要 :概述了前向网络结构目前的研究状况 .从隐层的本质特征出发 ,提出对样本数据采用模糊聚类分析法进行特征抽取 ,并综合考虑了样本的输入、输出信息 .实例分析说明此方法可行 ,较互相关分析法、奇异值分解法有更高的效率 .
关键词 :前向神经网络 ;隐层节点 ;特征抽取 ;模糊聚类 ;互相关 ;奇异值分解
中图分类号 :TP274⁺ .2 文献标识码 :A 文章编号 :1000-1980(2001)03-0017-04

前向网络(Feedforward Network)是目前应用最为广泛的神经网络模型之一 ,在水利水电工程中也取得了很多成果 ,但在应用神经网络解决实际问题时存在一个难点就是如何合理地确定隐层神经元节点数目 .规模过小的网络对训练样本的学习较为困难 ,而片面地追求网络的精度 ,选用较多的隐层节点数 ,易造成隐节点“ 冗余 ”现象 ,既增加了网络的训练时间 ,也使得网络模型的泛化性能(General Performance)降低 ,对非训练样本产生错误的识别结果 .为此 ,许多学者进行了创造性的研究 .主要有凑试法、增长法、修剪法 ,另外也有将进化方法、自适应方法和图论引入该领域的^[1] .增长法和修剪法分别又称构造法和析构法 ,是讨论得最为深入、集中的两种方法 .

增长法是从一个小规模网络结构开始 ,根据网络性能要求逐步增加结构复杂性 .Mezard^[2]第一个采用网络结构增长方法来解决 MFN 的局部极小问题 ,提出的 Tiling 算法一层一层地构造网络 ,通过在每层构造主节点和辅助节点以实现期望映射 : $R^n \rightarrow B^m = \{0,1\}^m$ 或 $\{-1,1\}^m$.Fahlmar^[3]提出了有名的 CC(Casade-Correlation)算法用于解决 BP 算法收敛速度慢的缺陷 .Littmann 等^[4]还提出了模块化增长方法 .修剪法开始时构造一个含有冗余节点的大规模网络结构 ,在训练过程中逐步删除那些不必要的节点或权值 .

复杂性调整方法在网络目标函数中引入表示网络结构复杂性的惩罚项 ,在优化过程中降低网络结构的复杂性实现网络的修剪 .目标函数一般定义如下形式 :

$$J(W) = E(W) + \lambda C(W)$$
 (1)

式中 : $E(W)$ ——网络误差项 ; $C(W)$ ——网络结构复杂性惩罚项 ; W ——权值集合 ; λ ——调整参数 . $C(W)$ 有多种定义方法 ,复杂性调整方法对 $C(W)$ 较为敏感^[5] .

汪家道等^[6]认为在网络训练时 ,各节点的输出可看作是随时间出现的序列 ,衡量隐层节点相似性的方法就是计算这些时间序列的互相关系数 ,计算式为

$$\rho_{o_i^{(k)} o_j^{(k)}} = \frac{\text{cov}(o_i^{(k)}, o_j^{(k)})}{\sqrt{D(o_i^{(k)})} \sqrt{D(o_j^{(k)})}}$$
 (2)

式中 : $\rho_{o_i^{(k)} o_j^{(k)}}$ —— $o_i^{(k)}$ 和 $o_j^{(k)}$ 的互相关系数 ; $o_i^{(k)}$ 、 $o_j^{(k)}$ ——第 k 层节点 i 与节点 j 的输出 ; $D(\cdot)$ ——方差 ; $\text{cov}(o_i^{(k)}, o_j^{(k)})$ ——协方差 ,

$$\text{cov}(o_i^{(k)}, o_j^{(k)}) = E\{[o_i^{(k)} - E(o_i^{(k)})][o_j^{(k)} - E(o_j^{(k)})]\}$$
 (3)

当 $\rho_{o_i^{(k)} o_j^{(k)}} = 0$ 时 ,表示 $o_i^{(k)}$ 和 $o_j^{(k)}$ 是不相关的 , $\rho_{o_i^{(k)} o_j^{(k)}}$ 越接近 1 ,说明 $o_i^{(k)}$ 和 $o_j^{(k)}$ 相关系数越大 ,若两神经元的互相关系数大于给定的阈值 D_{evc} ,则这两个节点进行合并 .

文献 [7,8] 则从矩阵奇异值分解的角度来压缩线性相关部分. 初选一个过大的网络结构, 训练至初步收敛, 将 m 个训练样本输入网络, 设第 i 隐层有 n 个节点, 其输出用 $m \times n$ 矩阵 A 来表示. 对 A 进行奇异值分解: $A = U \Sigma V^T$, 其中 $m \times m$ 矩阵 U 和 $n \times n$ 矩阵 V 是正交矩阵, Σ 是由 A 的奇异值构成的对角矩阵, $\Sigma = [\text{diag}\{s_1, \dots, s_p\} \ 0]$, $p = \min(m, n)$, $s_1 \geq s_2 \geq \dots \geq s_p > 0$, $s_1, s_2, \dots, s_p$ 称为矩阵 A 的奇异值. A 的 Frobenius 范数为

$$\|A\|_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} \tag{4}$$

通过证明可得

$$\|A\|_F = \sum_{i=1}^m \sum_{j=1}^n a_{ij}^2 = \sum_{i=1}^p s_i^2 \tag{5}$$

如果矩阵 A 的 p 个奇异值中有 q 个奇异值起主导作用, 也就是 $\sum_{i=1}^q s_i^2 / \sum_{i=1}^p s_i^2 > \lambda$. 那么可从 A 中选择一个 $m \times q$ 的 B 矩阵来逼近 A , B 中包含 A 的主要信息量, q 就是第 i 隐层所需的节点数. 逐次分析每个隐含层. 结构缩小后的网络重新初始化, 重新训练至收敛.

奇异值分解法对初始节点数和阈值有较大的依赖性, 由实例分析说明.

1 隐层节点数确定的模糊聚类分析法

网络的隐含层是输入模式的“内部表示”, 实际上就是对一类输入模式中所含的区别于其它类别的输入模式的特征进行抽取, 并将抽取出的特征传递给输出层, 因此, 中间层也称为特征抽取层. 隐含层节点数取决于输入模式特征 [9].

互相关系数法、奇异值分解法等均基于隐层节点输出值分析, 同时无论是增长法还是修剪法对初始网络结构隐层节点数无法确定, 要么太小, 要么很大, 这样要得到较优的结构需要很大的工作量. 虽然 Kolomogorov 定理 [10] 给出了 $2n+1$ 的理论计算值 (n 为输入层处理单元数), 但实际应用中效果并不理想.

本文提出一种基于模糊等价关系的特征抽取. 在输入模式中, 为了防止误把输入相近而期望输出值相差较大的样本归为一类, 考虑输入、输出的全部信息.

隐层节点数确定的模糊聚类分析法的具体步骤如下:

第 1 步建立新样本集.

设样本输入

$$\left. \begin{aligned} X &= [x_1 \ x_2 \ \dots \ x_n]^T \\ X_j &= [x_{j1} \ x_{j2} \ \dots \ x_{jn}]^T \end{aligned} \right\} \quad j = 1, 2, \dots, m \tag{6}$$

样本输出

$$\left. \begin{aligned} Y &= [y_1 \ y_2 \ \dots \ y_q]^T \\ Y_j &= [y_{j1} \ y_{j2} \ \dots \ y_{jq}]^T \end{aligned} \right\} \quad j = 1, 2, \dots, m \tag{7}$$

式中: x_{jk} ——第 j 个样本的第 k 个输入特征值; y_{jk} ——第 j 个样本的第 k 个输出特征值; m ——样本数目. 将样本输入、输出连接建立新样本集

$$\left. \begin{aligned} U &= [u_1 \ u_2 \ \dots \ u_n \ u_{n+1} \ u_{n+2} \ \dots \ u_{n+q}]^T \\ U_j &= [u_{j1} \ u_{j2} \ \dots \ u_{jn} \ u_{j(n+1)} \ u_{j(n+2)} \ \dots \ u_{j(n+q)}]^T \end{aligned} \right\} \quad j = 1, 2, \dots, m \tag{8}$$

第 2 步相似度量.

标定方法取夹角余弦法

$$r_{ij} = \frac{\left| \sum_{k=1}^{n+q} u_{ik} u_{jk} \right|}{\sqrt{\sum_{k=1}^{n+q} u_{ik}^2 \sum_{k=1}^{n+q} u_{jk}^2}} \tag{9}$$

相似性系数

$$r_{ij} \in [0, 1]$$

第 3 步构造模糊相似矩阵.

$$\tilde{R} = (r_{ij})_{m \times m} \tag{10}$$

第 4 步聚类分析.

采用直接聚类法 ,可以证明直接聚类法与传递闭包法是一致的^[11].取水平值 $\lambda \in [0,1]$,根据经验 λ 一般取 0.989 ~ 0.991.

若 $r_{ij} \geq \lambda$, U_i 与 U_j 归为一类 ,并将相连的类合并 .所谓两类相连 ,是指这两类的交非空 .最后的分类数即为前馈神经网络的隐节点数目 .

2 实例分析

为验证以上问题 ,以日本大阪湾海水测试的部分数据^[10]作为样本(如表 1) ,其输入变量为 x_1 ——水温 /℃ ; x_2 ——透明度 /m ; x_3 ——DO 浓度 /mg/L ; x_4 ——盐水浓度 /% ; x_5 ——过滤后的 COD 浓度 /mg/L ,输出变量为大阪湾的 COD 浓度 y .

表 1 海水测试数据
Table 1 Sea water test data

序号	x_1	x_2	x_3	x_4	x_5	y
1	12.5	2.5	7.4	29.3	1.8	2.9
2	9.3	2.4	9.4	28.9	2.2	2.8
3	9.5	2.5	9.4	30.1	2.1	2.7
4	12.9	2.5	9.1	29.2	2.1	4.3
5	16.0	2.1	10.0	25.8	2.9	4.9
6	19.7	1.6	10.0	22.9	3.0	5.9
7	22.8	3.0	5.0	22.2	2.0	3.2
8	28.2	1.0	10.0	25.9	3.3	8.4
9	24.6	2.8	6.4	30.7	2.1	4.3
10	22.2	2.3	5.9	30.1	2.6	3.6
11	17.9	2.5	6.7	27.0	2.0	3.0
12	14.3	4.0	7.5	28.1	1.5	3.1

根据实际问题取 5- P -1 的单隐层 BP 网络结构 ,采用模糊聚类分析法 , λ 取 0.99 ,计算结果样本分成两类 {6,8} ,{1,2,3,4,5,7,9,10,11,12} ,所以隐层节点数 $P=2$.

采用奇异值分解法 ,参数选取 :最大训练次数 5 000 ,误差指标 0.001 ,学习率 0.002 .5- P -1 的 BP 网络结构 ,隐层节点数 P 为 20,15,10,8,5,4 .训练至初步收敛 ,将其隐层输出进行奇异值分解 ,选阈值 0.99 ,由此得出隐层所需节点数(表 2) .可见奇异值分解法的结果并不唯一 ,与初选隐层节点数密切相关 .

表 2 奇异值分解结果
Table 2 Result of singular value decomposition

初选隐节点	奇异值	计算所需节点
20	11.985 3 ,1.490 5 ,0.929 7 ,0.578 9 ,0.382 4 ,0.047 5 ,0.000 1	2
15	10.295 4 ,0.326 8 ,0.179 4 ,0.011 9 ,0.000 1 ,	1
10	6.893 8 ,0.184 4 ,0.062 9	1
8	4.341 8 ,1.267 8 ,0.576 4 ,0.074 4 ,0.003 8	3
5	4.358 8 ,1.155 8	2
4	0.001 1	1

对初始网络结构 5- P -1 , P 和其它参数同奇异值分解法 ,分别将几种网络训练至收敛 ,用隐节点输出值的互相关分析法逐步进行隐节点自删除 ,计算的结论一致 ,网络隐层应有节点数为 2 .

从网络结构的确定到完成样本学习的全过程 ,模糊聚类分析法、奇异值分解法和互相关分析法的耗时分别为 39.4 s ,87.5 s ,93.2 s .可见虽然三种方法取得相似的结论 ,但模糊聚类分析法不需网络训练至收敛后判断调整 ,而是直接对样本数据进行分析 ,增加了特征抽取的透明度 ,学习效率也大大提高 .

参考文献 :

[1]何述东 ,瞿坦 ,黄献青 ,等 .多层前向神经网络结构的研究进展[J].控制理论与应用 ,1998 ,15(3):313~319.

- [2] Mzard M ,Nadal J P. Learning in feed-forward layered networks the tiling algorithm[J]. J Phys A Math Gen ,1989 22(8) :2191 ~ 2203 .
- [3] Fahlman S E ,Lebiere C. The cascade-correlation learning architecture[A]. Touretzky D S. Advances in Neural Information Processing System 1[C]. 1990. 524 ~ 532 .
- [4] Littmann E ,Ritter H. Cascade network architecture[A]. Proc of 1992 IJCNN[C]. 1992. 398 ~ 404 .
- [5] Weigend A S ,Rumelhart D E ,Huberman B A. Generalization by weight-elimination applied to currency exchange rate prediction[A]. Proc of 1991 IJCNN[C]. 1991. 837 ~ 841 .
- [6] 汪家道 ,孔宪梅 ,陈大融. 节点自删除神经网络及其在磨粒识别中的应用[J]. 清华大学学报(自然科学版) ,1998 38(4) :42 ~ 46 .
- [7] 陆系群 ,余英林. 设计前馈神经网络结构的一种新方法[J]. 华南理工大学学报(自然科学版) ,1998 26(1) :112 ~ 115 .
- [8] Psichogios D C ,Ungar L H. SVD-NET :An algorithm that automatically selects network structure[J]. IEEE Trans Neural Networks ,1994 , 5(3) :513 ~ 515 .
- [9] 王伟. 人工神经网络原理——入门与应用[M]. 北京 :北京航空航天大学出版社 ,1995. 67 ~ 70 .
- [10] 赵振宇 ,徐用懋. 模糊理论和神经网络的基础与应用[M]. 北京 :清华大学出版社 ,1996. 85 ~ 86 .
- [11] 张俊福 ,邓本让 ,朱玉仙 ,等. 应用模糊数学[M]. 北京 :地质出版社 ,1988. 233 ~ 236 .

Fuzzy Cluster Analysis Method for Determining the Number of Hidden Nodes of Feedforward Neural Networks

XU Shi-gang ,SUO Li-sheng

(College of Water Conservancy and Hydropower Engineering , Hohai Univ. , Nanjing 210098 ,China)

Abstract : In this paper , advances in architecture of feedforward networks are surveyed. Based on the essence of the hidden layer , a fuzzy cluster analysis method is put forward to extract the feature of sample data , moreover , the input and output information is comprehensively considered. An example has proved the feasibility of the fuzzy cluster analysis method , which is more efficient than the methods of interrelation analysis and singular value decomposition .

Key words : feedforward neural networks ; hidden nodes ; extraction of feature ; fuzzy cluster ; interrelation ; singular value decomposition