

BP 神经网络的优化及应用研究

吕砚山 赵正琦

(北京化工大学信息科学与技术学院, 北京 100029)

摘 要: 介绍对 BP 网络的结构和算法进一步优化的实验研究方法, 并给出了有关隐含层数和节点数选择、误差构成方式设计以及再学习策略引进的研究结果。

关键词: 人工神经网络; BP 算法; 液滴粒度分布

中图分类号: TQ 056.1

引 言

BP 神经网络在信息处理、智能控制等领域有着广阔的应用前景, 但在其应用上仍存在一些待解决的问题。例如, 在网络的结构方面隐含层节点数的选择尚无理论上的指导; 在算法方面标准 BP 算法存在局部最小问题未能解决, 以及迭代次数多、培训时间较长等^[1]。赖英旭等^[2]将 BP 网络应用于液滴粒度分布测量中, 并提出了一种通过实验确定单隐含层节点数的方法, 但对多隐含层的层数和节点数选择未作详细研究, 此外, 网络的算法也有待进一步优化。本文是在此基础上, 对上述问题作进一步探讨, 并结合将 BP 网络应用于液滴粒度分布测量和过热蒸汽流量测量的实例, 介绍对网络结构和算法优化的方法和结论。

1 网络结构的优化

BP 网络的输入层和输出层是与外界联系的接口, 这两层的节点数一般由应用对象的实际问题决定, 不能任意选取。可以改变的仅为隐含层的层数和各隐含层的节点数。

1.1 隐含层层数选择

首先探讨隐含层数不同对培训误差 S 的影响。以应用于液滴粒度分布测量中的 BP 网络为例, 具有单隐含层者, 网络各层节点数为 6-6-2; 再考虑一含双隐含层的结构, 各层节点数为 6-6-6-2 (图 1), 采用下列实验方法, 对两种网络进行培训; 用选定的同一培训样本, 按相同的培训方式 (成批处理修正误

差, 学习速率及动量因子不变) 及培训次数 (3 万次) 进行培训。比较其培训误差, 结果发现双隐含层网络的培训误差比单隐含层网络的培训误差小, 但是相差不是太大, 在培训次数 m 为 3 万次时, 单隐含层网络的培训误差为 0.34×10^{-5} , 双隐含层网络的培训误差为 0.11×10^{-5} 。此外, 与单隐含层网络相比, 双隐含层网络的培训误差下降速度较慢, 培训时间较长。

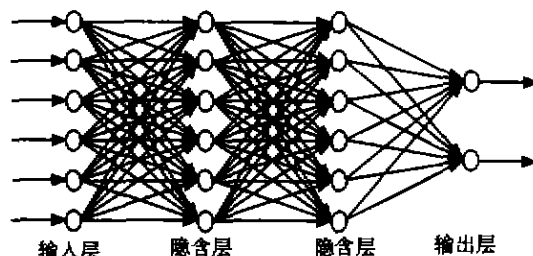


图 1 双隐含层 BP 网络

Fig. 1 Structure of two hidden layers BP network

为了考察隐含层层数对网络泛化能力的影响, 分别用上述培训好的网络对预测数据进行预测, 得到单隐含层和双隐含层网络的估计误差 S' 分别为 7.83×10^{-5} 和 2.16×10^{-5} 。这表明, 双隐含层网络的估计误差比单隐含层网络有所减小, 其改善程度比在培训误差方面要大, 说明双隐含层网络的泛化能力较高。可见, 在选择隐含层层数时要从网络精度和培训时间上综合考虑。对于较简单的映射关系, 在网络精度达到要求的情况下, 可以选择较少的隐含层层数, 以求加快培训; 对于较复杂的映射关系, 为了保证网络的精度, 则只能增加隐含层层数, 保证映射关系的正确实现。

1.2 隐含层节点数选择

隐含层节点数的选择与应用对象的具体问题有

收稿日期: 2000-06-26

第一作者: 男, 1936 年生, 教授

关,在不同的情况下可能选择不同的隐含层节点数。一般来说网络隐含层节点数太少,网络将不能建立复杂的映射关系,使网络培训不出来,或不能识别以前没有的样本,容错性差;节点数过多,又使网络学习时间过长,误差也不一定最小,因此 BP 网络应有一个最佳隐含层节点数。对此,鉴于尚无理论指导,也采用实验方法进行确定。以上述方法对输入和输出节点数不变、而隐含节点数改变的网络进行培训,比较其培训效果和泛化能力从而择优。在实验中初选隐节点数可按 $h = \sqrt{N+O} + a$ 式或 $h = \log_2 N$ 式确定大致范围^[3](式中 h 为隐含层节点数, N 为输入层节点数, O 为输出层节点数, a 为 0~10 之间的整数)。在建立网络时,对于隐含层节点的增减可采用以下二法:

构造法即先设置较少的隐含层节点数,若网络输出误差不符合设定的要求,则逐渐增加节点数,直到网络的误差不再有明显的减小为止;删除法即先给网络一个较大的隐含层节点数,若网络输出误差不符合设定的要求,则逐步删除隐含层节点,直至合适为止。

考虑到单隐含层网络的非线性映射能力较弱,对于相同的问题,为了达到预定的映射关系,隐含层节点要多一些,以增加网络的可调参数。反之,双隐含层网络的隐含层节点可以少一些。因此,单隐含层网络适用删除法;双隐含层网络适用构造法。

图 2 表示结构为 6- x -2 即具有单隐含层而隐节点不同时 BP 网络的培训误差。为了衡量对泛化能力的影响,表 1 中给出了具有双隐含层的 BP 网络的误差情况。可以看出,从 S 和 S' 综合考虑,双隐含层网络,当隐节点为(6,6)时,效果最好。此外,对单隐含层网络,节点数为 6 时效果最好^[2]。这说明无论单隐含层网络还是双隐含层网络,当隐含层

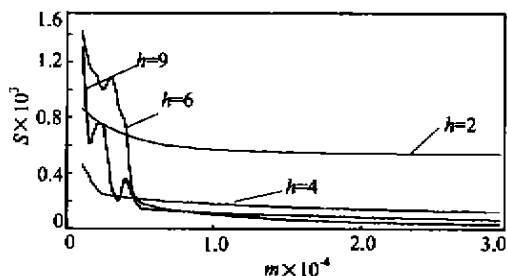


图 2 不同隐节点下的单隐含层网络培训误差

Fig.2 Training error of one hidden layer network in different hidden neurons

节点较少或较多时,网络的 S' 都比较大,只有在隐含层节点数适当时,网络的泛化能力最强。为了证明这一点,又将 BP 网络用于过热蒸汽流量测量中进行实验。为此,要建立输入节点数为 3(反应所测温度、压力、压差)、输出节点数为 1(要求给出流量),隐含层为单层节点数待优选的网络,以上述方法进行培训,用培训好的网络进行预测。其结果,也得到隐含节点过多或过少时 S' 较大,当隐节点数为 3(即网络结构为 3-3-1)时, S' 最小。此外,在培训过程中还发现,对于双隐含层网络,其两个隐含层节点数相同或接近时,网络的培训效果要比两个隐含层节点数相差较大时为好,如表 1 所示。

表 1 双隐含层下节点分布对培训误差和估计误差的影响
Table 1 Effect of neurons distribution of two hidden layers network on training error and predict error

h (第一层,第二层)	$S \times 10^{-3}$	$S' \times 10^{-4}$
(2,2)	2.854	152.41
(4,2)	0.216	
(4,4)	0.024	5.54
(4,8)	0.035	
(6,2)	0.165	
(6,6)	0.011	2.16
(6,10)	0.013	
(8,4)	0.032	
(8,8)	0.011	2.34
(8,12)	0.012	
(12,12)	0.011	3.40

由以上研究可知,网络隐含层结构的选择要从隐含层层数和隐含层节点数两方面综合考虑,并且还要与所要解决的实际问题联系起来。一般对于不太复杂的映射关系,应首先考虑应用单隐含层网络,用删除法找出一个优化的隐含层结构。对于比较复杂的映射关系,可以考虑用双隐含层网络或多隐含层的网络,用构造法确定网络的结构。针对上述具体应用对象,所得到的结果是:双隐含层网络的两隐含层节点数相近时网络的培训效果较好;单隐含层网络的隐含层节点数取靠近输入层节点数值时网络的泛化能力较强。

2 网络算法的优化

标准 BP 网络采用非线性梯度算法,不可避免地存在局部极小值问题,其迭代算法次数较多而学

习效率低,很长时间才能收敛,有必要进行优化。

2.1 误差修正方式

标准 BP 算法的误差修正方式基于一个样本,连接权的修改以一个样本为周期^[4],因此只反映了样本空间的局部性质。为了反映整个培训样本集的特性,采用了全部培训样本为一个修正周期,可以得到一种成批处理法,即在所有样本正向传播完后,计算它们的误差信号之和 $E = \sum_p E_p = \sum_p$

$\frac{1}{2} \sum_j (T_{pj} - O_{pj})^2$ (式中 T_{pj} 为样本 p 输出层节点 j 的期望输出; O_{pj} 为样本 p 输出层节点 j 的实际输出),根据此误差信号之和调整权值。由图 3 可以看出成批处理法的培训误差下降得比标准 BP 算法快,其收敛速度有较大的改善。而且成批处理法的培训误差也较小,在培训次数为 3 万次时,成批处理法 $S = 3.2 \times 10^{-5}$,标准 BP 算法 $S = 5.5 \times 10^{-5}$ 。为了检验成批处理法对网络泛化能力的影响,又用上述培训好的网络进行了预测,结果是:标准 BP 算法和成批处理法的 S' 分别为 9.76×10^{-4} 和 7.83×10^{-4} ,可见成批处理法的估计误差要小,这说明此法对网络的泛化能力有所改善。

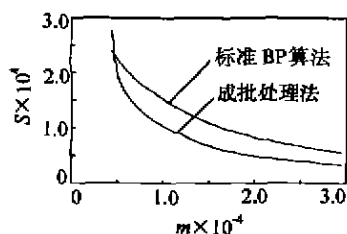


图 3 标准 BP 算法与成批处理法的培训误差

Fig.3 Training error of classic BP algorithm and batch processing algorithm

2.2 再学习策略

培训网络的目的是让它能够记住培训样本的特性,以便在预测时使用。为此,培训样本的选择必须

考虑可反映预测的实际情况。例如,在上述 6-6-2 网络的培训中,误差减小但不是最小值,说明可能陷于“局部最小”。为解决此问题,设法生成一个权值数据文件,然后重新调入此权值数据文件,用一个新的学习速率 η (由 0.6 变为 0.8) 使网络失去平衡。这相当于人为地给整个权值一个震荡,使其能够逃离局部最小束缚而误差进一步减小。图 4 显示了这种再学习策略的效果。

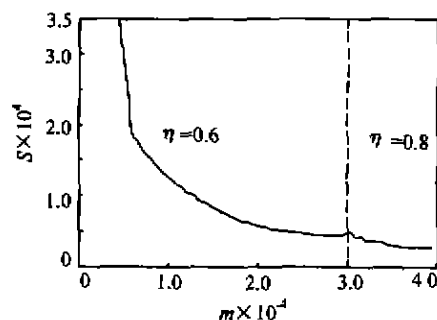


图 4 再学习策略的效果

Fig.4 Result of re-learning strategy

用上述培训好的优化网络对浙江大学化工系环化实验室直径为 300 mm 的旋流塔测得的脉冲速率函数^[5]数据进行了处理。结果表明,优化网络对未知样本的相对估计误差小于 1%,说明其泛化能力较强,满足了用户的要求。

参 考 文 献

- [1] 焦李成. 神经网络系统理论. 西安: 西安电子科技大学出版社, 1990. 36
- [2] 赖英旭, 吕砚山. BP 网络技术在液滴粒度分布中的应用研究. 北京化工大学学报, 1999, 26(2): 66~68, 92
- [3] 张立明. 人工神经网络的模型及其应用. 上海: 复旦大学出版社, 1993. 46
- [4] Graupe D. Principles of artificial neural networks. New York: World Scientific Publishing Co, 1997. 17
- [5] 吕砚山, 张进明. 气液两相流特征参数的电子测量. 山东: 山东科学技术出版社, 1992. 1

Optimization and application research of BP neural network

LU Yan-shan ZHAO Zheng-qi

(College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China)

Abstract: The further optimization of the structure and algorithm of BP network is introduced in this paper. It gives the research result of selection of the hidden layers and its neurons, design of error constitution pattern, introducing into the strategy of re-learning.

Key words: artificial neural network; BP algorithm; liquid-droplet size distribution