文章编号:1674-7070(2011)02-0164-04

人工神经网络泛化性能改进

赵远东[1]　胡为尧[1]

摘要

泛化能力是人工神经网络的重要特性.总结了增强神经网络泛化能力的几种方法,分析了各自的优缺点,并使用Matlab中提供的函数进行非线性函数的逼近,通过仿真与原函数拟合验证了泛化能力的提高.

关键词

神经网络;泛化能力;Matlab

中图分类号 TP391.9

文献标志码 A

0　引言

人工神经网络(Artificial Neural Network,ANN)是在对大脑的生理研究的基础上,用模拟生物神经元的某些基本功能元件(即人工神经元),按各种不同的联结方式组成的一个网络.模拟大脑的某些机制,实现某个方面的功能,可以用在模仿视觉、函数逼近、模式识别、分类和数据压缩等领域,是近年来人工智能计算的一个重要学科分支.

一个神经网络要有好的泛化能力,首先要保证网络的学习和训练要达到一定的精度要求.人们在神经网络的学习算法方面进行了大量的研究工作,提出了很多网络的改进算法,一般能够达到网络对学习精度的要求,但是当一个神经网络提高学习训练以后,达到了很高的精度要求,网络的泛化能力是否就很强了呢?人们在实际应用过程中经常会发现,即使网络的学习训练精度已经很高,其应用效果却不理想,就是网络在样本集中的映射结果很好,而应用到样本集外的数据时却产生了问题,也就是达到最小值后并不能保证网络对未知样本有较好的预测和推广能力.

目前,人们在神经网络的学习算法方面进行了大量的研究工作.然而,在如何提高神经网络的泛化能力方面,尚需做进一步的研究.

1　影响泛化能力的因素

神经网络的泛化能力,又叫推广能力,是指神经网络在训练完成以后输入其训练样本之外的新数据时获得正确输出的能力.它是ANN的一个属性,故又叫泛化特性.不管是什么类型的网络,不管它用于分类、逼近、推理还是其他问题,都存在一个泛化的问题.泛化特性在ANN的应用过程中表现出来,但由ANN的设计过程所决定.

人们已经发现许多影响泛化能力的因素,这些因素包括神经网络的结构复杂性、训练样本的数量和质量、初始权值、学习时间、目标规则的复杂性、对目标规则的先验知识等,但除了网络结构和训练样本数对泛化能力的影响已有一些定量的结果外,其余因素对泛化能力的影响还只有定性的解释,即影响神经网络泛化能力的主要因素是神经网络的结构复杂性和样本复杂性.

2　泛化方法

下面总结了几种改善神经网络泛化能力的方法,并通过举例验证.

收稿日期 2010-07-12

资助项目 国家重点基础研究发展计划项目(2009CB320501)

作者简介

赵远东,男,副教授,主要研究方向为人工智能与神经网络、计算机网络安全应用、计算机软件开发.ydzhao9@163.com

1　南京信息工程大学　信息与控制学院,南京,210044

2.1 结构设计

给定一组训练样本,存在同样本复杂性匹配的最小结构神经网络,该结构下的神经网络将有最好的泛化能力.这一原则同奥卡姆剃刀(Ocam's Razor)[1]原则是一致的,即在所有达到给定学习精度的神经网络中,结构越简单,泛化能力就越好.因此,当神经网络的结构复杂性与样本复杂性协调时,神经网络就会有较好的泛化能力.主要的结构设计方法有剪枝方法、构造方法、进化方法和信息论方法等[2],其基本思路都是通过调整神经网络的权值或隐节点数目,实现结构复杂性与样本复杂性的最佳匹配.

一般认为,网络拟合能力随着网络尺寸(包括网络隐层数目、各隐层节点、连接权值数目等)的增大而增大.文献[3]研究表明,隐层为 sigmoid 神经元、输出层为线性神经元的 2 层前馈网络,只要隐层神经元数足够多,可实现任意复杂的映射.因此,似乎可只简单地通过增加和减少隐层神经元的数目来调整网络的拟合能力.

现有的调整网络拟合能力以匹配任务复杂度的努力大概包括 3 个方面.其一是量化法,即想办法精确度量一个网络的与任务无关的拟合能力及一个任务的与网络无关的复杂度,并将二者进行比较以调整匹配[4];其二是逐步增大法,即先将网络初始化成非常简单,然后依据某种准则自动地逐步增加神经元个数以匹配;其三是自动修剪法[5],即将网络尺寸初始成足够大,然后依据某种准则逐步地修剪冗余权值或神经元.

刘妹琴等[6]构造基于 MDL 的目标函数,利用遗传算法优化 RBF 网络结构.

2.2 样本调整

给定一个假设空间 H,假设 h 属于 H,如果存在其他的假设 h' 属于 H,使得在训练样例上 h 的错误率比 h' 小,但在整个实例分布上 h' 比 h 的错误率小,那么就说假设 h 过度拟合训练数据,即为了得到一致假设而使假设变得过度复杂称为过拟合.想象某种学习算法产生了一个过拟合的分类器,这个分类器能够百分之百正确地分类样本数据(即再拿样本中的文档来给它,它绝对不会分错),但也就为了能够对样本完全正确地分类,使得它的构造如此精细复杂,规则如此严格,以至于任何与样本数据稍有不同的文档它全都认为不属于这个类别.过度拟合不利于提高神经网络的泛化能力.

训练样本的复杂度主要体现在质量和规模 2 个方面[7].质量是指训练样本的分布如何反映总体的真实分布,这是与采样过程相关的;规模则可以简单地理解为训练样本的数量.可见,训练样本的复杂度决定了训练集所包含的信息.样本质量指训练样本分布反映总体分布的程度,或者说整个训练样本集所提供的信息量.尽管样本质量对神经网络的泛化能力有相当大的影响,但定量分析样本质量对泛化能力的影响却是一个非常困难的课题.

一些预处理/滤波方法如主成分分析(PCA)、独立成分分析(ICA)则通过既降噪又降维,减小网络结构,从而避免学习过程中出现过拟合现象,改善网络泛化能力.武妍等[8-9]提出了一基于结果反馈的神经网络训练算法,将输入调整与权值调整 BP 算法结合起来,通过权值和输入矢量的双重调整来最小误差函数,该方法本质上是通过对输入样本集的模糊化处理避免学习过程中的过拟合,提高了泛化能力.

2.3 早期停止

早期停止法设计神经网络的主要思想是:最佳的泛化能力出现在训练误差的全局最小点出现之前,并给出了最佳泛化点出现的时间范围,从而从理论上证明了在神经网络训练过程中,存在最优的停止时间.这也说明,只要训练时间合适,较大的神经网络也会有好的泛化能力.

用早期停止法设计神经网络的关键是确定学习算法在何时停止学习.为了得到最优停止点,可以把所有样本数据分为训练数据和测试数据,并当测试误差达到最小时停止网络的训练.该方法实际上就是最简单的交叉测试(Cross Validation,CV)方法.

Matlab 中的 Early-Stopping 方法就是一个典型的例子.它在网络训练时从训练样本集中拿出一部分样本作为验证集(Validation Set)来控制训练的终止,当验证集上误差由减小开始增大时说明训练过程应该终止了.

3 3 种方法的优缺点

在 BP 网络中,设计神经网络的拓扑结构是非常重要的问题,它不仅对建立的神经网络模型的性能影响很大,而且可以导致"过拟合"现象.网络中隐层结点过少,则学习过程可能不收敛,但隐层结点数过多,则会长时间不收敛,还会由于过拟合,造成网络的容错性能下降,但是目前还没有一种理论能科学地确定隐含层结点数的方法.隐含层节点数不仅与输入/输出层的结点数有关,而且与需解决的问题的复杂程度

和转换函数的型式以及样本数据的特性等因素有关.

学习训练的样本对神经网络的泛化能力也有很大的影响. 如果数据样本数目太少,网络就很难学习到数据系列的规律,只能起到记忆数据的作用,有时甚至学习到错误的规律(相对于整个检验集). 当网络结构给定,即隐节点数 H 保持不变,若样本数 m 增加,网络泛化能力增强;反之,若样本数 m 减少,网络泛化能力减弱. 这只是一种最理想的状态下,并且是假设所有的样本数据都是类似的,而实际应用中,并不是所有的样本数据都是类似的.

早期停止的关键是如何确定训练停止时间. 如果仅根据训练误差曲线来确定停止时间,当训练样本的复杂性很高时,训练误差曲线可能不是递减的,这种技术并不适用.

4 验证

用早期停止方法来训练 BP 网络,使之逼近含有噪声的非线性函数 $y = \sin(2\pi p) + 0.1 \times \text{randn}(\text{size}(p))$. 建立只含有一个隐含层的 BP 神经网络,设计网络输入层神经元为 1,隐含层神经元为 20,输出层神经元为 1,则网络结构为 1—20—1. 隐含层选用双曲正切函数 tansig 传递函数,则输出层的选用 purelin 线性函数. 选取样本数据,则输入样本 $p = [-1:0.05:1]$;目标函数样本 $t = \sin(2\pi p) + 0.1 \times \text{randn}(\text{size}(p))$.

同样在验证样本中加入噪声,并且输入范围缩小为[-0.975,0.975],这样更符合实际样本,则验证样本为:输入样本 $P = [-0.975:0.05:0.975]$;目标函数样本 $T = \sin(2\pi p) + 0.1 \times \text{randn}(\text{size}(p))$. 用早期停止方法训练神经网络的程序如下:

```
p = [ -1:0.05:1];
t = sin(2 * pi * p) +0.1 * randn(size(p));
val.P = [ -0.975:0.05:0.975];
val.T = sin(2 * pi * val.P) +0.1 * randn(size(val.P));
net = newff([ -1 1],[20 1],{'tansig','purelin'},'traingdx');
net.trainParam.show = 25;
net.trainParam.epochs = 300;
net = init(net);
[net,tr] = train(net,p,t,[],[],val);
y1 = sim(net,p);
figure;
plot(p,t,'-',p,y,'--')
title('训练后的网络输出');
xlabel('输入样本')
ylabel('目标函数').
```

图 1 是采取一般训练方法的收敛曲线. 由图 1 知,网络训练至 300 步时停止,误差 0.007 455 3. 图 2 是早期停止法收敛曲线. 由图 2 可以看出网络训练至 131 步时提前停止,这是由于验证误差已经开始变大,此时网络误差为 0.009 097 53. 图 3 是早期停止方法拟合曲线. 由图 3 可以看出拟合效果比较满意,数据没有出现过拟合,早期停止法改善了神经网络的泛化能力.

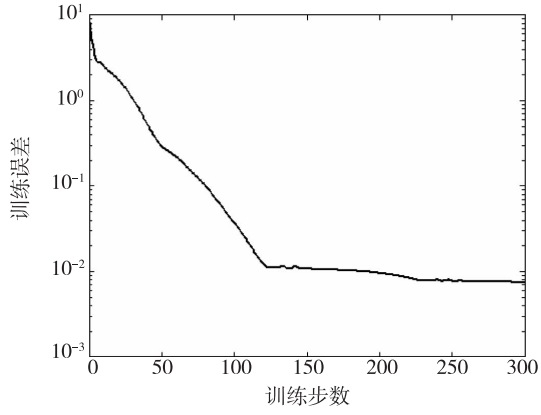


图 1 正常训练收敛曲线

Fig. 1 Normal training convergence curve

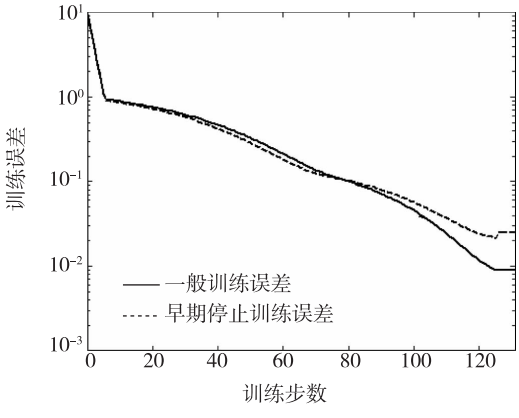


图 2 早期停止收敛曲线

Fig. 2 Early-stop convergence curve

5 结论

本文从结构设计、样本调整和早期停止 3 个方面介绍了几种提高神经网络泛化能力改进办法,分析总结了 3 种方法的特点,并以神经网络工具箱中提供的函数验证了早期停止方法,表明其网络的泛化性能有了较大的提高. 现有的网络泛化能力改善方法对一些简单问题是有效的,但是对较为复杂的问题效果还不明显,仍然存在着一些问题需要解决.

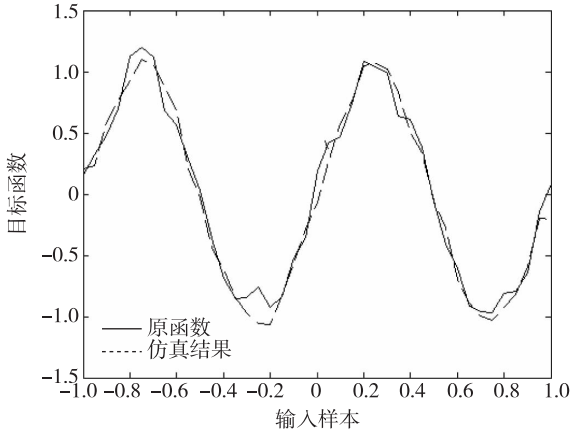


图 3 早期停止方法拟合结果

Fig. 3 Fitting result of early-stop method

参考文献

References

[1] 周洪斌. 粒子群优化算法应用研究[D]. 苏州:苏州大学计算机科学与技术学院,2009
ZHOU Hongbin. A study on applying particle swarm optimization algorithm[D]. Suzhou:School computer Science & Technology,Soochow University,2009

[2] 李爱军,罗四维,刘蕴辉,等. 信息理论框架下的神经网络构建[J]. 北京交通大学学报,2005,29(2):1-6
LI Aijun,LUO Siwei,LIU Yunhui,et al. Setting up neural network based on framework of information theory [J]. Journal of Beijing Jiaotong University,2005,29(2):1-6

[3] Sun J,Feng B,Xu W B. Particle swarm optimization with particles having quantum behavior[C]// Proceedings of 2004 Congress on Evolutionary Computation. Piscataway,NJ:IEEE Press,2004:325-331

[4] 黄山松,杜继宏,冯元琨. 前向神经网络的处理能力和推广性量度[J]. 清华大学学报:自然科学版,1999,39(7):579-581
HUANG Shansong,DU Jihong,FENG Yuankun. Capacity and generalization measurement of feedforward neural networks[J]. Journal of Tsinghua University:Science and Technology,1999,39(7):579-581

[5] 张鹤. 基于 SVM 的机动车流量数据分析与预测模型[D]. 沈阳:东北大学信息科学与工程学院,2008
ZHANG He. The analysis and predict model for traffic flow magnitude basised on SVM[D]. Shenyang:School Information Science and Engineering,Northeastern University,2008

[6] 刘妹琴,陈际达,蔡自兴. 基于 MDL 的 RBF 神经网络结构和参数的学习[J]. 小型微型计算机系统,2000,21(4):379-382
LIU Meiqin,CHEN Jida,CAI Zixing. Learning the architecture and parameters of RBF neural network based on MDL[J]. Mini-Micro Systems,2000,21(4):379-382

[7] Bao F,Pan Y H,Xu W B. A novel training algorithm for BP neural network[C]// Proceedings of the International Symposium on Distributed Computing and Application to Business,Engineering and Science. Shanghai:Shanghai University Press,2006:767-770

[8] 武妍,王守觉. 一种通过反馈提高神经网络学习性能的新算法[J]. 计算机研究与发展,2004,41(9):1488-1492
WU Yan,WANG Shoujue. A new algorithm to improve the learning performance of neural network through result-feedback[J]. Journal of Computer Research and Development,2004,41(9):1488-1492

[9] 武妍,王守觉. 基于模糊化输入和反转提高神经网络分类性能的方法[J]. 红外与毫米波学报,2005,24(1):15-18
WU Yan,WANG Shoujue. Method for improving classification performance of neural network based on fuzzy input and network inversion[J]. Journal of Infrared and Millimeter Waves,2005,24(1):15-18

Improve the generalization capability of artificial neural network

ZHAO Yuandong[1] HU Weiyao[1]

1 School of Information & Cybernetics,Nanjing University of Information Science & Technology,Nanjing 210044

Abstract The generalization capability is one most important performance of Artificial Neural Network(ANN). This paper discusses several methods on enhancement of ANN generalization capability,and analyzes their advantages and disadvantages. Functions provided by Matlab were employed to approximate the non-linear function,simulation and fitting with original fuction were carried out to verify the improvement of generalization capability.

Key words artificial neural networks;generalization capability;Matlab