

基于整合思想的神经网络泛化能力改进研究

黄宏涛

(广东财经职业学院信息管理系 广州 510420)

摘要 神经网络的泛化能力是影响其可用性的重要方面。本文在总结现有泛化能力改进措施的基础上,将各种独立的改进措施进行了整合,并融入神经网络的构造和训练过程,提出了基于整合思想的神经网络泛化能力改进措施,并针对其中的一类重要问题——确定训练时间进行了具体的数理分析,得出了一些有价值的结论。

关键词 神经网络,泛化,整合,训练时间

Research on the Generation Capacity of Neutral Network Based on Conformity Thoughts

HUANG Hong-Tao

(Information Management Department, Guangdong Finance and Economics College, Guangzhou 510420)

Abstract Generation capacity of neutral network is the important facet impacting on its usability. In this essay, many separate modified measures are conformed and the process of constitution and training of neutral network is fashioned. This essay makes an analysis on mathematical methods aiming at an important problem—conforming training time, and getting some valuable conclusions.

Keywords Neutral network, Generation, Conformity, Training time

1 引言

人工神经网络技术以其诸多显著优点,如非线性建模能力、强大的自学习能力、容错能力和自适应能力,受到青睐而被广泛地应用到科学研究和工程技术领域,并取得了很好的效果。神经网络本质上是一种工具,而我们使用这种工具一定是为了完成某项任务,即该工具一定要具有实用性。在神经网络的所有特性中,其泛化能力或者叫作推广能力是研究者关注的焦点之一。神经网络的泛化能力是指对于同一样本集中的非训练样本,网络仍能给出正确的输入输出关系的能力。由于在实际应用当中人们所关心的并不是网络对已知输入输出样本的拟合能力,而是对系统中已知输入未知输出(即非训练样本)的反映程度,因此使得网络的泛化能力成为网络应用成功与否的关键。网络的泛化能力可以理解为网络对系统的拟合能力。网络对系统的拟合程度越高,网络对系统中未训练样本的输入输出关系反映得越准确;反之,若网络的泛化能力强,即网络能正确反映样本集中未训练样本的输入输出关系,则网络对系统的拟合程度就高。在实际应用当中,由于系统的输入与输出之间的函数关系通常是未知的,人们只能通过对已知的有限样本的输入输出关系进行拟合,以此来代替对未知数据的拟合。因此可以说,没有泛化能力的神经网络没有任何使用价值。尽管目前已有不少学者研究了神经网络结构优化方法及提高神经网络泛化能力的方法,但基本上都是从某一个方面进行独立的改进和提高,没有最大程度地提高神经网络的泛化能力。本文在综合分析已有方法的基础上,提出了整合提高神经网络泛化能力的思想,即从多个方面入手,多管齐下,综合运用各种方法达到全面提高神经网络泛化能力的目的。

2 影响神经网络泛化能力的主要因素

影响神经网络泛化能力的因素有很多,但主要可以归结为以下三个方面:训练样本的数量和质量,网络结构如激发函

数、初始权值、目标规则的复杂性等,以及神经网络的训练时间。

2.1 用于训练的样本数据

成功训练一个神经网络有一个非常重要的前提条件,就是训练网络用的样本数据必须具有代表性。训练样本数据的选择和准备是成功构建一个运行良好的神经网络模型的第一步。要找到一个神经网络模拟,只有通过已知样本的训练来构造网络,在没有其他判定条件的情况下,显然只能认为对已知样本的拟合程度越高,网络对系统的拟合程度就越好,这使得训练样本的选取变得十分重要。训练样本越多,系统提供的信息越多,我们对系统的认识程度越高,网络经学习和训练后对系统的模拟程度也就越强。然而,在实际应用中,我们不可能无限制地增加训练样本,这种情况下就应尽量选取具有代表性的样本,即对系统特性刻画较好的样本。如何在有限的容量下选取具有代表性的、真实可靠的样本是决定网络泛化能力的关键,而这些既不是网络自身所造成的,也不是网络本身所能解决的。

此外,训练样本中的噪声数据也是影响神经网络泛化能力的一个重要因素。尤其是网络建立之初学习的前几个样本刚好受噪声影响较大,那么对于网络的健康生长更是十分不利。目前的神经网络增量学习类算法大都是采用每样本数据调整的训练方式,即在每一个样本进入网络之后,就进行参数调整。而样本中数据一般都是存在噪声的。每样本调整的训练方式,使网络更容易受到少数受噪声影响较严重的样本不良影响,降低算法效率和网络的泛化能力。

2.2 神经网络的运行机制

神经网络的运行机制主要包括网络结构(隐节点数,隐层数和隐节点的函数特性)和激发函数、权值确定等内容。目前的应用中,现有的神经网络算法过分追求网络对训练样本集中样本的训练误差达到最小,即通过增加网络的复杂性来降低训练误差,而忽略了由于网络复杂性的增加导致的过度学习现象,使网络的泛化能力下降。有研究表明,神经网络中各

隐层的隐节点数目不应多于样本个数,否则必有冗余的节点可以归并;各隐层的隐节点数目也应等于样本数,否则网络将成为插值网络。插值网络对已学习的样本的系统误差为零,但泛化能力较差。可是,当隐节点数目小于样本数时,如何确定合适的隐节点使网络的泛化能力达到最大,仍没有很好的结论。一般认为,隐节点数过小,网络结构过于简单,网络的泛化能力肯定不强;隐节点数过大,网络结构过于复杂,计算复杂性增加,而实用性降低。网络越复杂,所求解的非线性优化问题就越复杂,网络陷入局部极小点的可能性也就越大,这是造成复杂网络不如简单网络泛化能力强的主要原因。

对于神经网络来说,激发函数的选择也很重要,因为这直接关系到网络运行过程中参数权值的确定,并最终影响到运行结果。从数学优化的角度分析,神经网络的训练和构造最终都可归结为:通过选择激发函数来间接地确定网络的权值和阈值,使网络的实际输出与训练样本输出之间的误差达到最小,即求解非线性优化问题:

$$\min E_p = \frac{1}{2} \sum_p [t_k^p - \hat{y}_k(x^p)]^2$$

其中, t_k^p 是当网络输入为 x^p 时, k 输出单元的目标输出(教师输出)。

目前很多神经网络算法面临的一个重大问题就是由于激发函数的选择不当,致使算法不能最小化新数据提交给网络时网络实际产生的错误。换言之,我们期望的网络特性是其泛化到实际数据的能力。从神经网络训练的一般过程可以看出,训练一个网络的目的是使其最小化在训练样本数据上产生的错误。但当缺乏完备的、足够大的训练样本集合时,这一最小化目的很难实现。

2.3 神经网络的训练时间

神经网络的训练时间或者说学习时间也是影响其泛化能力的一个重要因素。过多的训练无疑会增加神经网络的训练时间,但更重要的是会使神经网络拟合数据中的噪声信号,产生所谓的过度学习,从而影响神经网络的泛化能力。有研究表明,在神经网络训练过程中,神经网络存在泛化误差的最小点。假定训练的目标是最小化误差函数,随着训练的进行,训练数据产生的误差自然下降,选择集数据产生的误差也同时下降。然而,如果前者在下降而后者不再下降,或者开始上升,这意味着网络开始过度适应数据,此时训练过程应中止。对于这种情况,通常建议减少隐藏的神经元个数/或者隐藏的网络层数,因为相对于给定的问题网络过分强大了。单隐层神经网络训练的泛化过程可分为三个阶段:在第一阶段,泛化误差单调下降;第二阶段的泛化动态较为复杂,但在这一阶段,泛化误差将达到最小点;在第三阶段,泛化误差又将单调上升。因此,最佳的泛化能力出现在训练误差的全局最小点出现之前,并且具有一定的时间范围。从另一个角度看,神经网络的训练应该是有“度”的,即存在一个最优的停止时间。因此,神经网络在训练过程中要严格控制训练时间,既不能训练不充分导致网络不成熟,又不能过度训练以致于降低效率和适用性,正所谓过犹不及。事实上,这两种情况都会降低神经网络的泛化能力。

3 基于整合思想的神经网络泛化能力改进措施

上面列举的几个因素都会从不同角度、不同程度地对神经网络的泛化能力发生影响,而且这些因素通常不是孤立的,而是同时发挥作用的,因此,要提高神经网络的泛化能力,我们应该统筹考虑所有这些因素,整合多种改良措施,形成提高泛化能力的合力。这就是基于整合思想的神经网络泛化能力改进的基本内涵。具体来说,应该从以下几个方面入手。

3.1 充分利用领域知识

神经网络毕竟是一种应用于某个领域的计算机或数学工具,因此,合理利用该领域的专业知识将会对提高神经网络的泛化能力产生重要影响。领域知识实际上是对所学习的目标规则所做的一些合理的限定。领域知识一般有两种表现方式:一种是以提示的形式出现;另一种是假定神经网络的激发函数权值服从某种先验分布。后一种方式在神经网络分析中引入贝叶斯方法,因而更易于人接受。使用领域知识,相当于对模型的参数加以限制,从而使模型更靠近“目标规则”所在的区域。事实上,加入领域知识相当于减小了神经网络函数类的维数,或者说增加了神经网络的训练样本。至于领域知识在神经网络中的使用,可以在学习过程中嵌入提示,或是在目标函数中增加一个权值限制项,来反映神经网络相邻输入之间的空间相关性,还可以将提示转化为虚拟样本,然后以虚拟样本和普通样本作为共同的训练样本集,完成对神经网络的训练。

3.2 选取合适的训练样本数据

在训练数据选择方面,由于训练数据是历史数据,是对历史事实的客观记载,一定程度上反映着某种潜在的关系。历史数据的产生有相应的自然的、社会的环境条件,因此历史数据隐含的关系依赖于当时的环境条件。如果环境条件变化很大,那么在过去成立的关系或许将来不再成立。因此,准备训练网络用的历史数据时,首先必须理清影响问题中数据产生的主要环境条件;然后以目前的环境条件作为基本参照来定位历史上相近或类似的环境条件,并收集相应的历史数据作为训练网络系统的样本数据。其次,准备的训练数据样本必须覆盖当前问题中出现的各种情况,使得训练数据是完备的。训练神经网络目的是最小化一个基于训练样本的全局误差,因此训练数据中含有不同类型的数据也是非常重要的。再次,样本数据的数量应该足够大,以便为学习算法提供丰富的学习信息,进而保证学习得到的网络系统是可信的、可靠的。

3.3 确定合理的初始权值

在合理确定初始权值方面,初始权值如何影响泛化性能和效率主要是由学习算法的动态来决定的。当使用梯度下降算法的时候,它趋向于收敛到接近初始条件的目标函数的最小点。在具有过多参数的网络里面,目标函数的最小点不是针对一个点,而是一个解平面。学习算法通常将权值调整到接近于初始条件的解平面的一个点。如果初始的权值较小,那么最后获得的权值也是小的,所以它的作用就类似于某种程度的权退化。为此,可以使用几个不同的初始权值范围分别对神经网络进行一定次数的训练,找到具有最小误差和最快学习速度的初始权值范围,然后使用优化的初始权值范围,继续训练神经网络,并且根据训练误差自动调整网络参数。

3.4 动态优化网络结构

在神经网络的训练过程中要进行动态控制和优化,综合使用各种控制网络复杂性的方法,如即增长法、剪枝法、正则化法等,将结构优化方法和正则化法结合起来共同来提高神经网络的泛化能力。或者采用逐渐修改网络结构,最终达到最优值的方法,每次修改后无须全部重新训练,而是利用前面得到的结果继续学习。在保证其稳定性和收敛性的条件下,运用领域知识自行设计神经元的激发函数,提高泛化性能。对函数逼近神经网络来说则可用权参数和隐节点数目来衡量。这里的函数逼近神经网络是指输入输出均取实值、隐单元取 Sigmoid 函数或 RBF 函数,而输出节点采用线性函数的前向网络,该网络常用于函数逼近。

4 通过交互对比合理确定训练时间

为获得较好的泛化能力,最关键的就是要合理确定训练时间,即在最合适的时候停止对神经网络的训练。为此,在神

神经网络算法中应当采用交互对比技术,结合最优停止准则,找出“最好”的神经网络结构,这里“最好”指的是网络具有较强的泛化能力,换言之,让神经网络在达到较好的泛化能力时即刻结束训练。交互对比技术将可用数据分为两个集合:训练集和测试集。训练集用来训练网络参数,测试集用来计算预测误差。然而,这种划分有可能引入偏差,因为只能用部分数据进行训练,另一部分用于测试。一个改进的措施是按几种不同方式、标准划分原始数据,进行多次训练和测试,这样就可以通过平均所有预测误差得出无偏的预测误差估计。

考虑一个隐层节点数为 N_h , 一个输出单元, 有 $N > N_h$ 个训练样本的神经网络。对于每一个隐层神经元, 都可以找到一个 N 维向量, 这个向量的第 i 个分量是这个隐单元对第 i 个训练样本的响应。既然网络输出是隐层输出的一个线性组合, 那么可能的输出向量存在于一个由 N_h 维基函数向量张成的 N 维向量空间当中。在 L^2 范数意义下, 最符合输出向量 t 的隐层输出向量无非是一个 t 到 N_h 维子空间的投影。这种情况下的误差平方和为: $E = t^T P^2 t$ 。此处, $P = I_N - \Phi A^{-1} \Phi^T$ 。其中, I_N 是一个 N 维单位矩阵, Φ 为基函数矩阵 ($(\Phi)_ij = \phi_j(x^i)$), $A = \Phi^T \Phi$ 。比如, 有 N 个数据点, $N - i$ 个用于训练, i 个用于测试。使用这种方法, 预测误差的方差估计是:

$$\delta^2 = \frac{t^T P i (\text{diag}(p))^{-2} P t}{P i}$$

其中, $P = I_N - \Phi A^{-1} \Phi^T$ 。

训练一个神经网络的目的之一是在过度学习和泛化能力之间寻找一个平衡点。上述交互对比的核心思想就是在训练

网络的过程中, 以一个独立的数据集为基准来检查算法的进展变化过程。网络在训练的样本数据上和选择集的样本数据上最初性能总是相同或者近似的。对于比较简单的神经网络系统, 如前馈神经网络不但对训练样本的模拟程度较好, 而且对未训练样本的拟合效果也比较理想, 即网络泛化能力较强。而对于比较复杂的系统, 要想使神经网络对系统的模拟程度提高, 在增加训练样本的同时也要增加网络的隐节点个数, 这样就使网络变得复杂。因此, 在训练网络时, 不必使训练误差达到尽可能的最小, 而给出一个适当的允许误差, 这样虽然网络对训练样本的拟合程度不是很好, 但网络的泛化能力增强。即先从一个神经元开始学习, 每次循环采用使网络产生最大泛化误差的输入作为新的隐层神经元。在具体实现中, 可以采用通过计算矩阵和学习输出向量之间的相对误差的办法来选择新加入的神经元, 然后用剩余的样本检查新网络的误差, 直到达到目标误差或最大神经元数为止。

参考文献

- 1 李杰, 韩正之. 神经网络的学习误差函数及泛化能力. 控制与决策, 2000, 15(1): 95~97
- 2 鲁子奕, 杨绿溪, 吴球. 提高前馈神经网络泛化能力的新方法. 电路与系统学报, 1997, 2(4): 69~72
- 3 张胜, 刘红星, 高教堂. 神经网络泛化特性改善方法. 计算机应用与软件, 2005(12): 12~14
- 4 张乃尧, 阎平凡, 李衍达. 神经网络与模糊控制. 北京: 清华大学出版社, 1998
- 5 武妍. 神经网络学习算法中的初始参数对泛化性能和效率的影响研究. 计算机工程与应用, 2002(23): 25~27, 31
- 6 An G. The effect of adding noise during back propagation training on a generalization performance. Neural Computation, 1996(8): 643~671

(上接第 227 页)

值求解的关节链序列。传统线性插值时, 关节链末端效应器沿运动路径相邻点的弧长是相等的; 而速度插值时, 末端效应器沿运动路径的各插值点之间的弧长随速度的变化而发生变化。由此可知, 确定的运动路径下, 角色对象的运动速度变化不同, 相同的对象模型可以产生与真实速度相映射的不同动作帧序列。

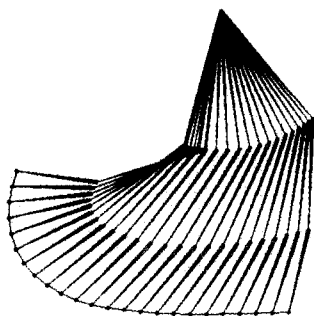


图 7 传统线性插值求解

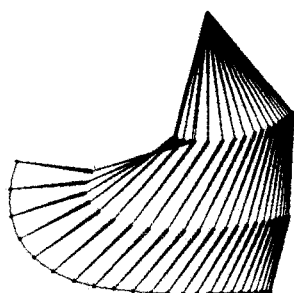


图 8 速度插值求解

结论 本文研究了隐曲线在角色行为动画中的应用。通过设置不同的运动路径和速度曲线, 速度插值技术可直观、快速地满足用户的具体要求创建不同的运动序列, 且运动复用性强。在未来的研究中, 可以通过将角色的风格控制并入逆向运动学技术, 可以得到更多真实感动画。

参考文献

- 1 Pullen K, Bregler C. Motion capture assisted animation: Texturing and synthesis. ACM Transactions on Graphics 2002, 21(3), 501~508
- 2 Gleicher M. Retargeting motion to new characters. In: ACM SIGGRAPH, Orlando, Florida, 1998. 33~42
- 3 Witkin A, Kass M. Spacetime constraints. Computer Graphics, 1988, 22(4): 159~168
- 4 Kovar L, Schreiner J, Gleicher M. Footskate cleanup for motion capture editing. In: ACM SIGGRAPH Symposium on Computer Animation, 2002. 97~104
- 5 Rose C F, Bodenheimer B, Cohen M F. Verbs and adverbs: multidimensional motion interpolation. IEEE Computer Graphics and Application, 1998, 18(5): 32~40
- 6 Kochanek D H U, Bartels R. Interpolating splines with local tension, continuity and bias control. Computer Graphics, 1984, 18(3): 245~254
- 7 Steketee S N, Badler N I. Parametric keyframe interpolation incorporating kinetic adjustment and phrasing control. In: ACM SIGGRAPH, San Francisco, 1985. 255~262
- 8 Witkin A, Popovic Z. Motion warping. In: SIGGRAPH'95, 105~108
- 9 Terra S C L, Metoyer R A. Performance timing for keyframe animation. In: Eurographic/ACM Symposium on Computer Animation, 2004. 253~258