

一种改进的神经网络增量学习算法

王 峰

(滁州学院电子信息工程系 滁州 239012)

摘 要 基于扩展 KALMAN 滤波器(Extended Kalman Filter)的神经网络是一类应用广泛的神经网络算法,但该算法在大数据量、抵抗噪声等方面还有相当的缺陷。本文从增量学习的角度出发,对扩展 KALMAN 滤波器算法进行了改进,同时借鉴周期算法的长处,引入部分增量训练机制(Partial incremental Training)和适当的隐层节点删减机制,使该算法在抵抗噪声等方面有了显著的提高。理论分析表明,该算法可以有效降低噪声数据的影响,提高神经网络算法的鲁棒性。

关键词 神经网络,滤波器 filter,算法,增量学习

A Modified Neural Network Increment Study Algorithm

WANG Zheng

(Department of Electronics and Information Engineering, Chuzhou University, Chuzhou 239012)

Abstract Neural network based on extended Kalman filter is a widely applicable neural network algorithm though it has problems in dealing with large data quantity and resisting noise. In this essay, the extended Kalman filter algorithm is modified from the point of incremental studying. The partial incremental training and proper obscure note cutting mechanism is introduced into the algorithm, considering the advantages of cycle algorithm, which makes it improved largely in resisting noise. Theoretical analysis manifests that the algorithm can reduce the impact of data noise effectively and improve the robustness of neural network algorithm.

Keywords Neural network, Filter, Algorithm, Increment studying

1 引言

目前的神经网络学习算法可以根据参数修正发生的时间分为两类:周期学习和增量学习。采用周期学习算法时,只在用全部训练样本进行了一次完整的学习之后才对网络参数进行调整。周期学习的过程是一个周期接着一个周期进行的,直到网络参数趋于稳定并且整个训练集上的平均输出误差收敛到某一个最小值时,学习才结束。采用增量学习算法时,则在每一个训练样本进入网络经过计算之后,都进行网络参数调整。具体来说,假如一个周期包含 N 个训练样本 $[x(1), y(1)], \dots, [x(N), y(N)]$ 。第一个样本 $[x(1), y(1)]$ 进入网络,计算出网络输出,然后,就运行学习算法对网络参数进行更新。重复这一过程直到对第 N 个训练样本学习完毕。前者更适合于静态的应用环境,即训练样本一次给出的情况,如分类问题等;后者则更适合于动态的应用环境,即不断有新的训练样本加入的情况,如动态预测问题等。RAN(Resource Allocating Network)算法就是一种应用比较广泛的增量学习神经网络算法。在通过调整隐层单元数目来反应要模拟的原函数的复杂度的思想基础上,RAN 学习算法的隐节点以“孤立性”作为引入以其为中心的基函数的标准,然后网络参数通过最小方差算法进行训练。当一个数据点远离现存的基函数中心,并且其网络输出与实际输出之间存在较大的误差时,则认为这个数据点是“孤立的”。如果没有新的隐节点需要加入时,启动最小方差算法对现存的隐层节点参数,包括中心、宽度和权值进行调整。

2 基于扩展 KALMAN 滤波器的神经网络算法简述

RAN 算法出现后,许多专家学者对该算法的思想进行了修正扩展,开展了卓有成效的研究。1993 年,两位波兰学者发表论文,从理论上讨论了采用增量学习算法的 RAN。他们提出使用扩展 KALMAN 滤波器替代最小方差法作为网络参数的训练算法,并且证明了该算法比 RAN 具有更快的收敛速度和更少的隐层单元数,这在增量学习的情况下是很有意义的。

扩展 KALMAN 滤波器的学习过程涉及引入新的隐层神经元以及网络参数调整。该算法从 0 隐节点的初始网络结构开始训练,样本数据进入网络时,算法从中选择一些作为隐节点中心增长网络结构。被选为新的隐节点中心,样本数据 (x_i, y_i) 需要满足以下两个条件: $\|x_i - c_{i'}^r\| > \epsilon_i$ 和 $\|e_i\| = \|y_i - \hat{y}(x_i)\| > e_{min}$ 。其中, $c_{i'}^r$ 为距离 x_i 最近的隐单元中心, ϵ_i 和 e_{min} 为事先选定的阈值。当一个新的隐节点加入到网络中后,其相关参数选为: $w_{N_h+1} = e_i$; $c_{N_h+1}^r = x_i$; $\sigma_{N_h+1}^r = \kappa \|x_i - c_{i'}^r\|$ 。其中, κ 为一个重叠因子,用来决定隐单元的响应宽度。在该算法中,设网络参数 $W_i = [w_0, w_1, c_1, \delta_1, \dots, w_{N_h}, c_{N_h}^r, \delta_{N_h}^r]$ 。 W_i 表示第 i 个样本数据进入网络后的参数状态。如果此样本数据不能满足成为新的隐节点中心的条件,则算法根据 W_{i-1} 调整网络参数 W_i : $W_i = W_{i-1} + K_i e_i$ 。由于扩展 KALMAN 滤波器的快速收敛性,有可能导致求得的模型对未来数据的适应性差。为此,其他研究者建议加入随机项: $P_{i(x \times z)} = [I - K_i B_i^T] P_{i-1} + q_0 I$, 其中, q_0 为随机步长,用

来决定在梯度方向上随机取值的大小。当一个新的隐层节点引入网络后,就增大了 P_i 的维度,此时就需要在 P_{i-1} 中加入新的行和列。 p_0 是付给参数的初始值的不确定性的估计,在这里就是样本数据 x_i 和 y_i 的协方差。 z_1 是由于新增隐节点引进的新增参数的个数, $z_1 = m + n + 1$ 。算法初始时, $\epsilon_i = \epsilon_{\max}, \epsilon_{\max}$ 为输入空间中非 0 可能的最大跨度。在随后的网络训练成长过程中, ϵ_i 以指数速度减小,逐渐衰减,直到 $\epsilon_{\min}$ 为止。 ϵ_i 以指数速度减小的策略,使 RBF 网络从具有较大宽度的较光滑基函数开始学习,然后随着更多隐节点加入网络,很快就可以更精确地逼近目标值。

无论 RANEKF 还是 RAN,因为采用了增量学习的策略,都能得到较预先设定隐层神经元数目算法更精简的网络结构。事实上,样本数据存在噪声,神经网络在训练过程中,将不可避免地引入以噪声较严重的样本数据点为中心的隐层神经元。另外,在一些特殊的应用条件下,网络的建模对象原问题本身也可能随着时间发生模型本身的细微变化,这时,网络当中部分原有隐节点就有可能变得不活跃。这些节点一方面是多余的,增加了额外的计算负担;另一方面,更重要的是,使网络产生过拟和现象,降低了网络的泛化能力。然而,这两个算法中,都只有引进新的隐层单元的机制,而没有判断、删减不活跃甚至是错误引入的隐层节点的机制,一个隐节点一旦被引入就无法再从算法中删除。本文通过研究发现,扩展 KALMAN 滤波器算法的隐层节点部分存在作用相似节点并存的现象,久而久之容易形成节点冗余从而降低训练效率。此外,由于该算法本身比较复杂,在对计算效率有较高要求的应用场合,算法的实用性可能会受到影响。

3 改进后的扩展 KALMAN 滤波器神经网络增量学习算法

传统的径向基神经网络都是采用输入层-隐层-隐层-输出层全连接的。本文认为,在很多情况下,这种全连接方式并不都是必要的。在没有任何关于原问题的先验知识的情况下,采用全连接确实是一种简化问题处理的有效手段。但是这种简化也有可能带来不必要的计算效率降低和网络冗余。与最终输出结果无关的输入分量会给网络带来成倍增长的多余隐节点,带来极大的计算负担。如果能够识别和发现这样的输入分量并做针对性的网络结构调整,那么最终网络结构将可以成倍精简,极大提高算法效率和更好地逼近原问题的隐含规律。对每一个新进训练样本,该算法由于只是依据网络输出误差和最近隐单元距离来判断新样本的“孤立性”,采用简单的方差对比作为增加隐层单元的标准,当一个被噪声污染较严重的样本进入网络,就很可能被选为新隐节点的中心。因为这个新节点与最近现有隐单元的距离也许就大于规定的阈值,同时网络输出误差也较大。这样的结果就是,网络为了拟和噪声,而错误地引入了新的隐节点。为了减少噪声对网络的影响,并使网络隐单元增长平缓,本文对基于扩展 kalman 滤波器的神经网络算法进行了改进,提出了一种改进后的基于扩展 kalman 滤波器神经网络的增量学习算法——Incremental Learning Algorithm Based on Extended Kalman Filter Neural Network(ILA)。

3.1 部分增量训练方法

如前所述,扩展 Kalman 滤波器算法较易于受单个受污染样本数据的影响。这是一个增量学习算法普遍存在的问题。相对于增量学习算法,周期算法在这一点上有一定优越

性。因为周期算法在学习过所有训练样本才进行参数调整,多个样本一同训练可以达到平滑噪声影响的作用。因此,本文考虑是否可以将周期算法的这种优势引入增量学习算法中,以提高本文算法的适用性。具体来说,就是通过几个样本同时学习,延缓参数更新时机,降低个别污染较重的样本数据对网络的影响,从而提高网络的泛化能力。本文将此方法称为部分增量训练法(Partial incremental Training),具体做法是:设置一个样本缓冲区,一个新样本进入网络,按以下步骤处理:①先用前述的 RAN 相关判断准则检验该样本数据是否符合作为新增隐节点中心的条件,符合条件转②,不符合转③;②如果缓冲区为空,则以此节点为新隐层中心加入新隐单元,处理完毕;如果缓冲区不空,则用缓冲区中全部现有样本数据训练网络,更新参数,并清空缓冲区,转①;③将该样本数据放入缓冲区,如果放入后,缓冲区满,则用缓冲区中全部现有样本数据训练网络,更新参数,并清空缓冲区。缓冲区的大小决定了部分增量的大小,可以根据实际问题试探确定。

3.2 隐节点删减策略

针对扩展 Kalman 滤波器算法不能发现和处理网络中不活跃、作用很小的隐节点的不足,本文在算法中引入了适当的隐节点删减策略。事实上,此前已有专家学者提出了在周期学习算法中加入删减隐节点的思想。在他们的方法中,每个周期都要检查各隐节点的权值,那些权值小于某个阈值的隐节点就要被删除。受此方法启发,本文对已有的删减思想进行了改进,不仅考虑了隐节点权值的大小,还考虑了隐节点的输出值的情况。

在径向基神经网络中,一个隐节点对于网络输出的贡献取决于三个因素:连接权值 w_{ik} 、输入数据与隐节点距离 $\|x_i - c_k\|$ 、隐函数宽度 σ_k 。仅仅考虑权值是不充分的。试想,如果宽度 σ_k 很小或距离 $\|x_i - c_k\|$ 很大,隐节点的输出也会变小。对于一个多输出的径向基函数神经网络($n > 1$), σ_{ik}^l ($l = 1, \dots, n$) 是 k 隐节点输出向量 $O_k = [o_{ik}^1, \dots, o_{ik}^l, \dots, o_{ik}^n]^T$ 的第 l 个分量, O_k 可以用下式计算: $O_k = w_{ik} \phi(x_i)$ 。

为了判断一个隐节点是否应该被删除,所有隐单元的输出向量的每一个分量都应当被考虑。如果一个隐节点的输出向量 O_k 一段时间之内都十分不显著,就应当考虑删除它。这个时间长度可以用学习样本数据的个数来衡量,比如,设当前在学习的样本为 t ,时间长度 np 就表示学习从 $t - np + 1$ 到 t 本共 np 个样本这段时间。要判断输出向量 O_k 是否不显著,需要考察它的每一个分量是否都不显著。考虑到不同隐节点输出的不一致性,仅仅比较输出绝对值是不合适的。这里应该先对隐层输出向量 O_k 进行标准化处理: $r_k = [r_{ik}^1, \dots, r_{ik}^l, \dots, r_{ik}^n]$, $r_{ik}^l = \|o_{ik}^l\| / \|o_{i, \max}^l\|$ 。其中, $\|o_{i, \max}^l\|$ 是输入 i 样本时,所有隐单元对 l 输出单元输出的最大绝对值,即: $\|o_{i, \max}^l\| = \max(\|o_{i1}^l\|, \dots, \|o_{ik}^l\|, \dots, \|o_{inp}^l\|)$, ($l = 1, \dots, p$)。

本文认为,这种标准化的办法并不完全合理,因为我们要研究的是一个隐节点对于输出层的贡献,以 $\|o_{i, \max}^l\|$ 为标准化的基准,不能直接反映隐单元对输出单元输出的贡献率。直接使用网络输出作为基准可能更为合适,则上式相应修改为: $r_{ik}^l = \|o_{ik}^l\| / \|\hat{y}_l(x^i)\|$ 。如果在连续 np 个训练样本期间,标准化输出向量 r_k 的所有分量都小于一个预先选定的阈值,就说明 k 隐节点对于整个网络输出的贡献已经十分不显著,应该将其从网络中删除。据此,隐结点删减策略可以总结如下:

(下转第 238 页)

绘制能力。另一个可以进行的工作是进行交互技术的研究,设计一套更加有效的交互工具,对多维函数提供更加灵活的交互操作。

最后,本文提出的算法可以方便地对体数据进行交互式绘制,更有效地了解感兴趣区域内部结构的空间关系,使得体绘制更具实用价值。

参考文献

- Nicoletti G M. Advances in Direct Volume Rendering for Visualizing Large 3D Data Sets from Scientific and Medical Applications. In: Proceedings of the 5th Biannual, World Automation Congress, 2002. 245~250
- Lorensen W E, Cline H E. Marching Cubes: a High Resolution 3D Surface Construction Algorithm. In: Proceedings of ACM SIGGRAPH'87, 1987. 163~169
- Pfister H, Lorensen B, Bajaj C, Kindlmann G, Schroeder W, Avila L S, Raghu K M, Machiraju R, Lee J. The Transfer Function Bake-off. IEEE Computer Graphics and Applications, 2001. 16~22
- 黄汉青,唐泽圣.应用于传递函数设定的交互式体绘制工具[J].计算机学报,2005,28(6): 1062~1067
- Kindlmann G, Durkin J W. Semi-Automatic Generation of Transfer Functions for Direct Volume Rendering. In: Proceedings of IEEE Volume Visualization Symposium, 1998. 79~86
- 王文成,孙汉秋,吴恩华.加速体绘制技术[J].计算机辅助设计与图形学学报,2002,14(9): 880~886
- Kniss J, Kindlmann G, Hansen C. Multidimensional Transfer Functions for Interactive Volume Rendering. IEEE Transactions on Visualization and Computer Graphics, 2002,8(3): 270~285
- Cabral B, Cam N, Foran J. Accelerated Volume Rendering and Tomographic Reconstruction using Texture Mapping Hardware. In: Proceedings of ACM Symposium on Volume Visualization, 1994. 91~98
- He T, Hong L, Kaufman A, Pfister H. Generation of Transfer Functions with Stochastic Search Techniques. In: Proceedings of IEEE Visualization, 1996. 227~234
- Marks J, Andalman B, Beardsley, Pfister H, et al. Design Galleries: A General Approach to Setting Parameters for Computer Graphics and Animation. ACM Computer Graphics (SIGGRAPH'97), 1997. 389~400
- König A, Gröller E. Mastering Transfer Function Specification by Using VolumePro Technology. In: Spring Conference on Computer Graphics, 2001, 17: 279~286
- Levoy M. Display of Surfaces from Volume Data. IEEE Computer Graphics and Applications, 1988, 8(3): 29~37
- Kniss J, Kindlmann G, Hansen C. Interactive Volume Rendering using Multi-dimensional Transfer Functions and Direct Manipulation Widgets. In: Proc. of IEEE Visualization, 2001. 255~262
- Kindlmann G, Whitaker R, Tasdizen T, Möller T. Curvature-based Transfer Functions for Volume Rendering: Methods and Applications. In: Proc. of IEEE Visualization, 2003. 513~520
- Lum E B, Ma K L. Lighting Transfer Functions using Gradient Aligned Sampling. In: Proc. of IEEE Visualization, 2004. 289~296
- Westermann R, Ertl T. Efficiently Using Graphics Hardware in Volume Rendering Applications. ACM Computer Graphics (SIGGRAPH'98 Proceedings), 1998. 169~179
- Engel K, Kraus M, Ertl T. High-Quality Pre-Integrated Volume Rendering Using Hardware-Accelerated Pixel Shading. In: Proceedings of Eurographics/SIGGRAPH Workshop Graphics Hardware, 2001. 9~16
- Jie Zheng, Hongbing Ji. High Quality Per-Pixel Shading in Texture-Based Volume Rendering. In: Proceedings of the International Conference on Neural Networks and Brain (ICNN&B2005), 2005. 1258~1261
- nVidia Developer Relation Site. Available online at: <http://developer.nvidia.com>

(上接第 178 页)

对每一个输入样本 (x_i, y_i) 按以下步骤处理:①计算所有隐节点的输出向量 O_k ($k=1, \dots, N_k$);②计算网络输出 $\hat{y}(x_i)=[\hat{y}_1(x_i), \dots, \hat{y}_l(x_i), \dots, \hat{y}_n(x_i)]^T$;③为每一个隐节点计算归一化输出向量 r_k ;④删掉在连续 np 个样本输入期间归一化输出向量 r_k 的每个分量都小于阈值 δ 的隐节点;⑤如果采用扩展 Kalman 滤波器算法,根据简化的网络结构调整有关矩阵的维度;⑥修剪隐层-输出层连接。

根据以上删减算法,只有那些在连续 np 个样本输入期间对每个输出节点的输出都不显著的隐节点才会被删减。在多输出的结果中,可能存在这样一些隐节点,它们对部分输出单元的输出很不显著,但是对其他一些输出单元有较为显著的输出。对于这样的隐单元上述算法不作处理。实际上,隐层-输出层连接的不显著也隐含了关于背景问题的知识。它可能说明这个隐层节点与其相连接的输出节点的相关性不大,从而反映出某些输入样本与特定输出节点的相关性不大。据此,本文进一步提出了修剪隐层-输出层连接的机制,并与前述隐节点删减算法合并为:

对每一个输入样本 (x_i, y_i) 按以下步骤处理:①计算所有隐节点的输出向量 O_k ($k=1, \dots, N_k$);②计算网络输出 $\hat{y}(x_i)=[\hat{y}_1(x_i), \dots, \hat{y}_l(x_i), \dots, \hat{y}_n(x_i)]^T$;③为每一个隐节点计算归一化输出向量 r_k ;④修剪在连续 np 个样本输入期间都小于阈值的 δ 归一化输出向量的分量 r_{ik} 所对应的隐层-输出层连接,即令 $w=0$;⑤删掉在连续 np 个样本输入期间归一化输出向量 r_k 的每个分量都小于阈值 δ 的隐节点;⑥如果采用扩展 Kalman 滤波器算法,根据简化的网络结构调整

有关矩阵的维度。经过多余隐单元修剪可以提高网络泛化能力,而通过隐层-输出层连接修剪,可以简化网络结构,提高神经网络的学习效率。

结束语 本文的算法中尝试通过适当的隐节点删减策略合并这些相似隐单元,给出了修剪隐层-输出层连接的机制,将它合并到了隐层节点删减的算法当中,以进一步简化网络结构,更好地拟合原问题。相信这一机制将对多输出的神经网络训练产生有益的影响。但是另一方面,采用 ILA 算法虽然能够取得几乎最小的网络结构,但是在算法执行前有许多预选参数要设定,关于预选参数的设定是一个尚需深入研究的问题。此外,ILA 算法的许多技术方法是基于经验和直觉的,缺乏正式的理论证明,这也是许多神经网络研究的共同问题。这些问题都需要在本文的后续工作中加以深入研究。

参考文献

- 阎平凡,张长水.人工神经网络与模拟进化计算.北京:清华大学出版社,2000
- Lu Y W, Sunderarajan N, Saratchandran P. A Sequential Learning Scheme for Function Approximation Using Minimal Radial Basis Function Networks. Neural Computation, 1997(9): 461~478
- Bors G, Pitas I. Median Radial Basis Function Neural Network. IEEE Transactions on Neural Networks, 1996, 7(6): 1351~1364
- 肖国强,张为群.连续学习混沌神经网络的研究.计算机科学,2004(4): 12~16
- 王晓丹,郑春颖,吴崇明.一种新的 SVM 对等增量学习算法.计算机应用,2006(10): 45~48