

基于注意力模型的混合学习算法*

杨博⁺, 苏小红, 王亚东

(哈尔滨工业大学 计算机科学与技术学院, 黑龙江 哈尔滨 150001)

A Hybrid Algorithm Based on Attention Model

YANG Bo⁺, SU Xiao-Hong, WANG Ya-Dong

(College of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

+ Corresponding author: Phn: +86-451-86413142, E-mail: boyang621@126.com, <http://www.hit.edu.cn>

Received 2003-11-04; Accepted 2005-01-06

Yang B, Su XH, Wang YD. A hybrid algorithm based on attention model. *Journal of Software*, 2005,16(6): 1073-1080. DOI: 10.1360/jos161073

Abstract: A hybrid algorithm based on attention model (HAAM) is proposed to speed up the training of back-propagation neural networks and improve the performances. The algorithm combines the genetic algorithm with the BP algorithm based on magnified error signal. The key to this algorithm lies in the partition of the BP training process into many chips with each chip trained by the BP algorithm. The chips in the same iteration are optimized by the GA operators, and those in different iterations constitute the whole training. Therefore, the HAAM obtains the ability of searching the global optimum solution relying on these operations, and it is easy to be parallelly processed. The simulation experiments show that this algorithm can effectively avoid failure training caused by randomizing the initial weights and thresholds, and solve the slow convergence problem resulted from the Flat-Spots when the error signal becomes too small. Moreover, this algorithm improves the generalization of BP network by improving the training precision instead of adding hidden neurons.

Key words: back-propagation algorithm; artificial neural network; attention model; genetic algorithm; Flat-Spots; local optimum

摘要: 为了解决传统 BP(back-propagation)算法收敛速度慢,训练得到的网络性能较差的问题,在借鉴生理学中“选择性注意力模型”的基础上,将遗传算法与误差放大的 BP 学习算法进行了有机的融合,提出了基于注意力模型的快速混合学习算法.该算法的核心在于将单独的 BP 训练过程划分为许多小的切片,并对每个切片进行误差放大的训练和竞争淘汰机制的选择.通过发现收敛速率较快的个体和过滤陷入局部极值的个体,来保证网络训练的成功率和实现快速向全局最优区域逼近的目的.仿真结果表明,该算法有效地解决了传统 BP 算法中由于初始权值的随机性造成的训练失败问题,并能有效解决饱和区域引起的后期训练缓慢问题,在不增加网络隐层节点数的情况下,显著地提高了网络的收敛精度和泛化能力.这将使神经网络在众多实际的分类问题上具有更广泛的应用前景.

* Supported by the National Natural Science Foundation of China under Grant No.60273083 (国家自然科学基金)

作者简介: 杨博(1978—),男,贵州贵阳人,博士生,主要研究领域为人工智能,模式识别,多智能体系统;苏小红(1966—),女,博士,副教授,主要研究领域为机器学习,色彩匹配,计算机图形学;王亚东(1964—),男,教授,主要研究领域为知识工程,专家系统,生物信息技术.

关键词: BP 算法;人工神经网络;注意力模型;遗传算法;饱和区域;局部极值

中图法分类号: TP18 文献标识码: A

BP(back-propagation)算法作为一种神经网络训练方法,由于其理论依据坚实,推导严谨,物理概念清晰,目前仍是应用最为广泛的神经网络模型之一^[1-3]。但该算法也同时存在着学习时收敛速度慢,所得到的网络性能较差的缺点,因此,针对快速学习算法的提出和改进也一直是神经计算的重要研究内容。到目前为止,已有很多研究者进行了这方面的研究,并取得了大量成果。如附加动量的 BP 算法、可变学习速率的 BP 算法、共轭梯度法、拟牛顿法、Prime Offset 算法和 Delta-bar-Delta 学习规则等。但这些改进算法仍然存在着由于初始权值产生的随机性造成的易陷入局部极小点的问题,以及陷入饱和区域^[4,5]造成的后期学习缓慢、训练精度难以提高的问题。

通常,训练精度的提高主要是依靠增加隐含层的节点以及训练时间来获得^[6],如果隐含层节点数设置过多,虽然可能会提高训练样本的精度,但这时的网络并不能真实反映对象的规律,而只是将对象作为高维映射的一个特例,更重要的是,网络复杂度的增加以及对学习样本的过度拟合,将导致神经网络泛化能力的下降。统计学习理论(statistical learning theory)的研究成果表明,要获得一个泛化性能较好的学习系统,降低 VC 维是一个可行的途径,而这是可以通过尽量使用简单的网络结构对训练集进行训练来获得的^[3,7]。这说明通过增加隐节点来提高网络训练精度的方式并不十分合理。因此,在不增加网络结构复杂度的情况下,研究如何提高网络的性能具有重要的意义。

遗传算法通过模拟自然界的进化过程来迭代产生适于解决问题的优化解,其搜索机制中的隐含并行性使得搜索过程能不断向全局最优解进行逼近。遗传算法的并行性表现在:(1) 内在并行性(inherent parallelism),它可以在分布式并行处理环境中各自进行独立种群的演化计算;(2) 隐含并行性(implicit parallelism),由于演化计算采用种群方式组织搜索,从而它可以同时搜索解空间内的多个区域,并相互交换信息,这种搜索方式使得虽然每次只执行与种群规模 N 成比例的计算,而实质上已进行了大约 $O(N^3)$ 次有效搜索^[6,7]。这使得演化计算能以较少的计算获得较大的收益。

神经网络与遗传算法的结合被认为是再现智能行为的一个很有希望的途径^[1],通常采用一些全局优化算法来解决神经网络训练中的局部极小问题。为了解决神经网络训练收敛缓慢、易陷入局部极小点的问题,目前的研究主要集中在两个方面:一是对神经网络初始权值进行优化^[8-10],而遗传算法本身也存在早熟收敛问题,因此这些方法仍然不能保证后继的网络训练不会陷入局部极小区域;二是采用进化的神经网络方法(evolutionary neural networks,简称 ENN)^[3,6,11-13],完全用遗传算法替代 BP 学习,以避免梯度下降法的缺陷。但是,由于遗传算法本身微调能力较弱的特性使得进化神经网络需要一个较大范围的初始权值区域,由此造成的复杂度升高使得训练出的神经网络泛化能力并不理想。同时,遗传算法的种群计算常常使得其训练的时间开销比 BP 算法的时间开销大得多。

为了更好地发挥遗传算法的全局寻优能力和 BP 算法的局部微调能力,本文基于注意力模型提出一种混合算法——Hybrid Algorithm based on Attention Model(HAAM)来解决网络的高精度训练问题,该方法可以在不增加网络结构复杂度的情况下,快速提高网络的学习精度,同时又不会降低网络的泛化能力。

该混合学习算法的基本思想为:通过对神经网络的训练过程进行切片,并将每一切片嵌入到遗传算法中进行演化计算和误差放大训练,达到对权值组合空间进行并行爬坡的目的,减少以往基于梯度下降法中单一爬坡造成的训练失败和陷入局部极值的发生率。一方面,该方法可以避免由于随机产生初始权值将网络训练引入局部极小区域,确保训练有一个较高的成功率;另一方面,通过误差放大训练和“自然选择”的竞争机制快速发现更适于解决问题的网络个体,去除陷入局部极小区域而收敛缓慢的个体,以使具有简单网络结构的神经网络训练能以较高的精度快速收敛,同时又具有较好的泛化能力。

根据统计学习理论,在网络结构复杂度即 VC 置信度相同和不降低网络泛化能力的前提下,训练精度越高即经验风险越小的网络,越能保证期望风险较小,而且还能使网络训练收敛到全局极小点的概率大为增加。本文通过对三异或问题、噪声模式分类问题和两个实际的数据分类问题进行仿真模拟实验,验证了这种新的混合算法在不增加网络结构复杂度和降低网络泛化能力的情况下进行高精度训练的有效性和较好的泛化性能。

1 视觉选择性注意力模型

集中注意力是人脑合理而巧妙地通过感官有选择地接受和处理外来信息的方式.在这种注意力高度集中的状态下,人的感官和大脑对所集中注意力部位信息接收的灵敏度和信息处理速度都大大增强^[2,14].通常会对比平时容易被忽略的细微变化做出巨大反应.这可以看作是人的感官对外界刺激的一种放大效应,使被注意事件的刺激信号得到加强,从而能够更有效地对其进行响应和处理.

为了探求注意力机制,Treisman 等人于 1986 年提出了一个富有成效的模型——视觉注意力的特征综合理论^[15,16],如图 1 所示.该理论认为,注意力的选择机制是由两个相互独立的、分级的官能共同组成的两个阶段:前注意阶段(pre-attentive stage)和注意阶段(attentive stage).在前注意阶段中主要是对视野区域中的目标特征作无容量限制的并行特征探测.而在注意力阶段,一个目标(或有限的几个目标)被称为“聚光灯(spotlight)”或“变焦透镜(zoom lens)”的过程进行特征放大处理.当目标从第 1 阶段进入第 2 阶段时,这些目标就被认为是被选择的目标^[15].

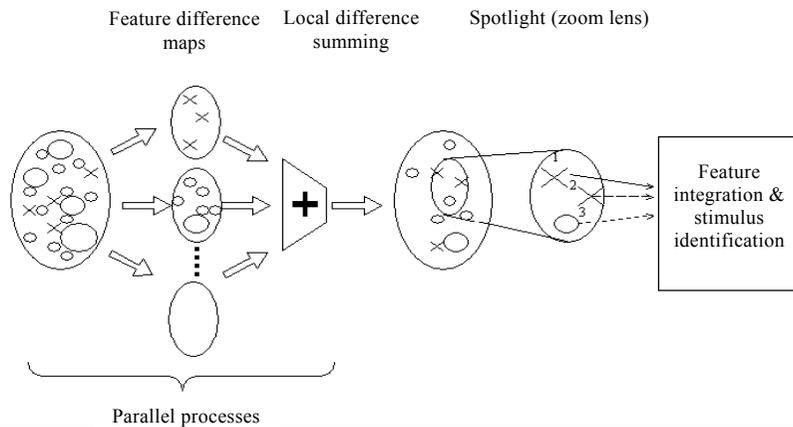


Fig.1 A model of visual selective attention

图 1 一个视觉选择性注意力模型

目前,大量的心理物理、神经生理学实验及计算机仿真工作已经给出注意系统一些基本框架.例如,心理物理学领域广泛采用“聚光灯”隐喻(spotlight metaphor),认为视觉选择性注意可以增强视场中被注意区域范围内的信息处理效率,而将聚光灯照射范围之外所有区域的信息过滤掉^[14,17];以及被认为更贴切些的“变焦透镜”隐喻(zoom lens metaphor),认为注意力能对所关注的局部场景进行尺度缩放^[14,15].

这样就使得注意力部位对被注意对象的各种特征信息有更加细致的观察.犹如借用高倍放大镜来加强被注意对象的各种细节,使之更容易被观测和被控制.

2 基于注意力模型的混合学习算法

根据上述研究成果的启发,模仿人脑在处理被注意事件的过程中,对被注意对象的各种特征信息进行加强处理,本文提出一种新的基于注意力模型的混合学习算法 HAAM.该算法的基本思想如下:

首先将神经网络的训练过程(最大迭代次数 T)划分为 ρ 个切片,每个切片长度 $L \in \{x | 0 < x \leq 100, x \in \mathbb{Z}\}$,并对遗传算法中的初始种群进行误差放大的切片训练,通过“优胜劣汰”的演化机制,实现对陷入局部极值的(权值,阈值)组合进行过滤,并产生新的种群以进行下一轮的切片训练.

其中的种群对应于注意力模型中当前的视野区域,误差放大的切片训练犹如注意力阶段中进行的“变焦透镜”处理过程,实现对细节信息的加强处理.

HAAM 算法的伪码描述.

Begin

确定网络训练的最大迭代次数 T ,切片因子 ρ ,切片长度 L ,求解终止条件(精度 c),种群规模 N ,交叉概率 P_c ,

变异概率 P_m ;

```

 $A_0$ =问题空间中随机产生  $N$  个个体;
While Termination_Contitions=False
  For ( $i=0; i<N; i++$ )
    If (Fitness[ $i$ ]>Fitness[Elite])
      Elite $\leftarrow A_k[i]$ ;
    End If
     $A_k$ =Select( $A_{k-1}$ )
    Crossing_over ( $A_k$ );
    Mutate( $A_k$ );
    If ( $\rho>0$ )
      For ( $i=0; i<N; i++$ )
        Train( $i,L$ );
         $\rho\leftarrow\rho-1$ ;
      End If
    End While;
End.

```

对于待求解问题中包含大量局部极小值的情况,可将 ρ 设置较大,以利用 GA 的全局寻优能力帮助 BP 网络训练跳出局部极值区域.反之,则可以将 ρ 设置较小,以充分利用 BP 训练的局部微调能力快速逼近极值点.

其中,误差放大训练 Train(i,L)采用改进的 BP 算法,通过将标准 BP 算法中误差项的计算式(1)替换为自适应误差放大的计算式(2)来获得:

$$\delta_j = \begin{cases} o'_j(t_j - o_j), & \text{如果 } j \text{ 是输出层单元} \\ o'_j \sum_k w_{kj} \delta_k, & \text{如果 } j \text{ 是隐层单元} \end{cases} \quad (1)$$

$$\delta_j = \begin{cases} (t_j - o_j)M_j, & \text{如果 } j \text{ 是输出层单元} \\ o'_j \sum_k w_{kj} \delta_k, & \text{如果 } j \text{ 是隐层单元} \end{cases} \quad (2)$$

M_j 是聚光灯增强系数:

$$M_j = \frac{1}{o'_j + E(j)} \quad (3)$$

式(3)的作用是根据当前误差函数的计算结果和误差的梯度特征自适应调整增强系数 M_j ,使得输出误差项这一细节信息得到加强,以便更有效地对当前网络状态进行学习.

由式(3)可以看出,当网络训练到高精度时,输出层的误差项将会被放大到对权值修正有明显影响的范围.在传统的 BP 算法中,由于受到饱和区域的影响,网络状态越接近目标状态,正比于激活函数导数的权值修正量越小,目的是避免修正过量引起的学习震荡现象,但同时也使得越接近目标解,学习的效率越低.当权值修正量与待修正权值的数量级相差较大时,权值的修正对网络的输出值几乎不会造成任何影响^[4,5],这是网络训练的学习速率逐渐下降直至停滞的直接原因.

HAAM 算法通过式(2)来调整网络权值修正的有效范围,使得网络参数的后期学习仍然维持在一个较高水平,即让修正值的大小始终接近于待修正权值的量级水平.而算法中采用的精英保留策略,对当前搜索到的最优解进行独立记录或保存,又减轻了震荡带来的负面影响,并能保证算法最终能收敛到全局最优解.

如图 2 所示,实心三角形为待识别的模式,由于整个模式空间中有许多干扰模式存在,虽然遗传算法可以去除部分干扰模式,但对于很相似的“局部极小”解很难进一步提高,通过模式特征的放大可以将相似模式有效地区分开.

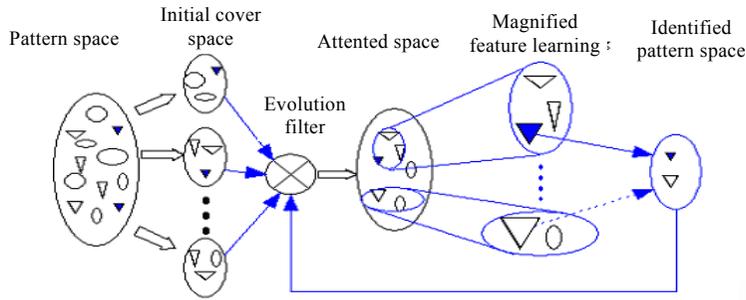


Fig.2 HAAM working flow

图2 基于注意力模型的混合学习算法工作流程

与以往混合学习算法的不同之处在于,该算法中的遗传演化机制和网络训练机制是一个交叉进行的过程,通过切片式的阶段性网络训练,及时对收敛缓慢的(权值,阈值)组合进行过滤,大大提高了网络训练的成功率和收敛速度;同时,对切片进行的误差自适应放大训练,还起到了抑制饱和和区域干扰收敛精度的作用,使网络训练能够快速搜索到较高精度的解.这与演化算法中“两阶段算法”^[18]所使用的小生境技术不同,HAAM 算法中没有采用对初始群体进行再划分的小生境技术,而是对整个神经网络的训练过程(总迭代次数 T)进行了切分,以便及时发现和处理训练过程中陷入局部极值和饱和区域的情况.

从对算法的描述还可以看出,通过对切片的适当控制,用遗传算法对神经网络初始权值进行优化的方法和采用进化神经网络(ENN)的方法实际上都是 HAAM 方法的一个特例,或者说,HAAM 方法是这两种方法的一个推广.当 $\rho=0$ 时,整个神经网络的(权值,阈值)组合完全依靠 GA 进行优化,即退化成为进化神经网络的方法;当 $\rho=1$ 时,在 GA 算法迭代完成后再启动 Train 算子,即退化成对神经网络初始权值进行优化的方法;当 $\rho>1, N=1, P_c=P_m=0$ 时,HAAM 就退化成为一个基于误差放大的 BP 学习算法,其 BP 迭代次数为 $\rho \times L$.

3 仿真模拟实验分析

将提出的 HAAM 算法应用于解决几个典型问题的实验上:三异或问题、最小趣味编码问题(minimum interesting coding problem,简称 MICP)^[19,20]以及 Soybean-small dataset,Iris dataset 的实际分类问题.遗传迭代中的适应值采用训练样本的累加误差平方和,以求得神经网络输出误差最小时的一组神经网络的权值组合^[6].

3.1 3层BP网络权值的学习

分别用加入动量项(momentum)的 BP 算法、Prime Offset 算法、进化神经网络(ENN)和基于注意力模型的混合学习算法(HAAM)训练一个 3 层 BP 网络的权值.其中,学习系数为 0.8,惯性系数为 0.7, $N=10, \rho=100, L=1, P_c=0.5, P_m=0.08$,offset 为 0.1.

网络结构为 3-3-1,训练样本取自 3 异或问题的神经网络,共需训练 12 个连接权和 4 个阈值.终止条件是训练次数 T 达到 100 代,成功训练的标准为:所有样本的累加误差平方和小于 0.1.

分别对每个算法进行 50 次单独训练,各算法成功收敛的平均结果如图 3 所示.其中,各算法每一次训练的初始权值在 $[-10,10]$ 上随机产生,算法的性能比较见表 1.

比较图 3 中的收敛曲线,可知 HAAM 算法的学习速度最快,能很快逼近全局最优解;除 HAAM 算法以外,其余算法在 100 次的训练周期内,都无法收敛到该全局最优解.它们在达到一定精度时就陷入 sigmoid 饱和区域,从而使学习率急剧下降,即使是针对饱和区域缺陷进行一定改进的 Prime Offset 算法的效果也不是那么理想.可见,HAAM 算法能有效解决饱和区域抑制精度提高的问题,学习精度有了较大的提高.

从表 1 可以看出,HAAM 算法在 100 次迭代内就能达到 $1e-30$ 以上的精度值,这是其他几种网络学习算法所不能达到的.同时从整个算法的学习速度来看,本算法也要比启发式变异算法^[11](heuristic mutation algorithm,简称 HMA)的效率高很多,文献[11]中的 HMA 算法在训练后期需要近 1000 次迭代,才能将精度提高一个数量级,而本算法在后期训练的实验中只需 100 次迭代就能使精度提高十几个数量级.同时,HAAM 算法用于网络训练

的成功率也较其他几种算法要高,有效地解决了传统 BP 算法中由于初始权值的随机性造成的训练失败问题.

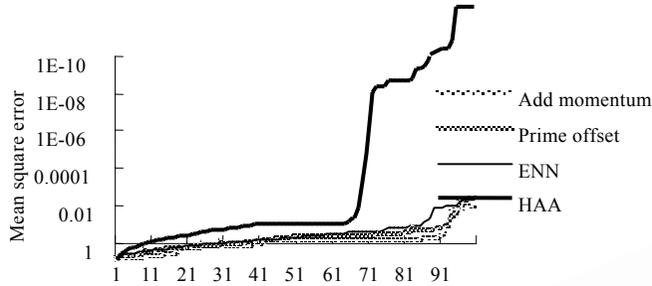


Fig.3 Average result for 50 times experiments

图 3 50 次实验的平均结果图

Table 1 Comparison on performances of the four algorithms

表 1 4 种算法性能比较

	HAAM	ENN	Prime offset	Momentum
Best mean square error	6.16e-33	1.63e-7	5.10e-4	2.62e-3
Times of convergence to the best result	48	30	27	4
Success rate (%)	96	60	54	8

3.2 噪声模式分类的学习

用一个噪声模式分类的学习任务来检验 HAAM 算法用于快速学习的有效性和泛化能力.该问题是一个有噪声输入的最小趣味编码问题^[19,20](minimum interesting coding problem,简称 MICP),问题描述见表 2,在 4 位输入中,前两位是噪声,与输出模式无关,后两位与输出的关系是二进制格雷码编码(Gray coding)关系.

Table 2 Minimum interesting coding problem

表 2 最小趣味编码问题

	1	2	3	4	5	6	7	8
Input	0000	1100	1001	1101	0010	0110	0011	1011
Output	00	00	01	01	11	11	10	10

	9	10	11	12	13	14	15	16
Input	0100	1000	0001	0101	1010	1110	0111	1111
Output	00	00	01	01	11	11	10	10

实验用模式集中的前 8 个作为训练集,后 8 个作为测试集,以检测学习的泛化能力和抗噪声性能.网络结构为 4-4-2,使用标准 BP 算法,Prime Offset 算法和 HAAM 算法进行性能比较,其中,学习系数为 0.8, $N=10, \rho_1=10, \rho_2=50, L=10, P_c=0.5, P_m=0.08$,offset 为 0.1,初始权值在[-1,1]上随机产生.各算法 20 次实验的平均结果见表 3.

从表 3 中可以看出, HAAM 算法无论是在收敛速度、收敛精度还是在误判率上都比参与比较的其他算法优越得多.实验同时表明,Prime Offset 算法虽然在训练集上比标准 BP 算法有较高的收敛精度,但其泛化性能却有所下降,出现了学习过拟合现象;而 HAAM 算法有效地避免了这种情况的发生,它通过嵌入在遗传算法中的误差放大训练算子,增强了遗传算法的微调能力;也通过遗传算法的并行优化能力,使网络训练不易陷入局部极小点,在提高神经网络收敛精度的同时不仅没有降低网络的泛化能力和抗噪能力,而且使训练得到的网络具有更好的泛化性能.

Table 3 Comparison of learning on MICP

表 3 MICP 训练结果比较

Comparison items	Criteria rule		Training time (ms)	Test error	Fault rate (%)
	Criteria error	Max epochs			
BP	1e-3	10000	262.5	2.18e-01	10
Prime offset	1e-4	10000	181.2	3.39e-01	11.88
HAAM	1e-6	10x10	45.2	8.79e-04	0

3.3 对实际分类问题的学习

用 HAAM 算法和进化神经网络(ENN)方法分别对 Soybean-small dataset 和 Iris dataset 进行数据分类识别的学习,这两个数据集来自于美国加州大学 Irvine 分校的机器学习数据库,被广泛地用于对机器学习中各种算法的性能测试^[5,21,22].数据集特征如下:

Soybean-small dataset:4 类,35 个特征,47 个数据(第 1~3 类各 10 个数据,第 4 类 17 个数据).

Iris dataset:3 类,4 个特征,150 个数据(每一类 50 个数据).

实验采用的网络结构分别为 35-4-4,4-4-3;并使用 k -Fold Cross Validation^[21,23]($k=10$)方法进行验证,将数据集划分为 10 个集合,每次用其中的 9 个作为训练集,剩余的作为测试集,结果取 10 次实验的平均值.实验结果见表 4.

Table 4 Comparison on practical cluster problem with real datasets

表 4 实际分类问题训练结果比较

	ENN		HAAM	
	Soybean-small	Iris	Soybean-small	Iris
Fault rate (%)	4	2.67	0	2.0
Training time (s)	35.234	23.585	1.192	16.613

从表 4 可以看出,在相同环境、相同规模下,HAAM 算法较 ENN 方法具有更强的学习能力,训练时间大大减少,并且能更有效地提高网络性能.

从实验算例的结果可以看出,HAAM 算法通过对神经网络训练过程的适当切分,利用 GA 的全局寻优能力和竞争筛选机制及时发现和处理了陷入局部极值和饱和区域的情况,减少了传统训练算法在局部极值区域反复迭代造成的大量时间开销,从而有效地提高了学习效率.而算法中的误差放大训练过程,有效地利用了 BP 训练的局部微调能力,弥补了 GA 微调能力不足的缺陷.因此,该算法更适用于解空间中局部极值点较多、易受饱和和区域影响的神经网络学习问题.

4 结束语

本文基于注意力模型,提出一种新的快速神经网络混合学习算法,仿真实验结果表明,在使用 Sigmoid 型激励函数的人工神经网络中,这种基于注意力模型的混合学习算法(HAAM)比传统的 BP 算法及其变形具有更好的学习效果,不仅能够提高训练的成功率,而且收敛速度快,可以达到很高的学习精度.HAAM 算法通过在遗传算法中嵌入误差放大的训练算子,增强了算法的微调能力,使其能够快速向全局最优解收敛;并通过遗传算法的并行优化能力,使其中的神经网络训练不易陷入局部极小点,提高了训练的成功率.同时,由于在提高精度时没有增加网络结构的复杂度,对权值取值范围的复杂度也进行了适当控制,因此得到的网络也具有较好的泛化能力.

由于遗传算法本身的并行特性,使得 HAAM 算法也具有进行分布式并行求解的能力,通过切片的并行求解,为神经网络的分布式训练提供了一种可行的设计方法,因此,进一步的研究工作是通过深入研究注意力模型中的并行机理,寻找注意力选择机制和大规模神经网络并行学习的最佳结合点,以便给大规模神经网络的学习提供一条并行训练的有效途径.

References:

- [1] Lu RQ. Knowledge Engineering & Knowledge Science of the Century. Beijing: Tsinghua University Press, 2001 (in Chinese).
- [2] Jin F. The Intelligence Basis of Neural Computing: Theory & Method. Chengdu: Southwest Jiaotong University Press, 2000 (in Chinese).
- [3] Yan PF, Zhang CS. Artificial Neural Network and Simulating-Evolution Computation. Beijing: Tsinghua University Press, 2000 (in Chinese).
- [4] Balakrishnan K, Honavar V. Improving convergence of back propagation by handling flat-spots in the output layer. In: Alek-Sander I, Taylor J, eds. Proc. of the 2nd Int'l Conf. on Artificial Neural Networks. Neural Networks, 1992,5:139-144.

- [5] Parekh R, Balakrishnan K, Honavor V. An empirical comparison of flat-spot elimination techniques in back-propagation networks. In: Proc. of the 3rd Workshop on Neural Networks-WNN'92. 1992,55-60.
- [6] Cong S. Neural Network, Fuzzy System & the Applications of Motion Control. Hefei: Press of University of Science and Technology of China, 2001 (in Chinese).
- [7] Lu RQ. Knowledge Science & Computing Science. Beijing: Tsinghua University Press, 2003 (in Chinese).
- [8] Yang B, Su XH, Wang YD. BP Neural network optimization based on an improved genetic algorithm. In: Yeung DS, Wang XZ, eds. Proc. of the 1st Int'l Conf. on Machine Learning and Cybernetics. Beijing: IEEE Press, 2002. 64-68.
- [9] Zhai YF, Li HY, Liu HB, Yuan XM. Method for optimizing initial weights of ANNs by Gas. Journal of Jilin University (Engineering and Technology Edition), 2003,33(2):45-50 (in Chinese with English abstract).
- [10] Yin JX, Chen SY, Qiu J. Neural network prediction model and its application based on GA and BP. Journal of Dalian University of Technology, 2002,42(5):594-598 (in Chinese with English abstract).
- [11] Zheng ZJ, Zheng SQ. Study on a mutation operator in evolving neural networks. Journal of Software, 2002,13(4):726-731 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/13/726.pdf>
- [12] Tao Q, Cao JD, Sun DM, Fang TJ. A dynamic genetic algorithm based on the neural network with constraints. Journal of Software, 2001,12(3):462-467 (in Chinese with English abstract).
- [13] Yang GJ, Cui PY, Li LL. Applying and realizing of genetic algorithm in neural networks control. Journal of System Simulation, 2001,13(5):567-570 (in Chinese with English abstract).
- [14] Crick F. The Astonishing Hypothesis: The Scientific Search for the Soul. New York: Charles Scribner's Sons. 1994.
- [15] Treisman A. Features and objects in visual processing. Scientific American, 1986,(254):114-124.
- [16] 2003. <http://www.ahisee.com/content/attentlessay.html#TOC1>
- [17] Guo AK. Computing Neuroscience. Shanghai: Shanghai Science and Technology Press, 2000.
- [18] Li Y, Kang Z. A two-level subspace evolutionary algorithm for solving multi-modal function optimization problems. Wuhan University Journal of Natural Sciences, 2003,8(1B):249-252.
- [19] Schaffer JD, Caruana RA, Eshelman LJ. Using genetic search to exploit the emergent behaviour of neural networks. Physica D 42, 1990,42:244-248.
- [20] Zhang J, Liu KS, Wang XF. Immune modulated symbiotic evolution in neural network design. Journal of Computer Research & Development, 2000,37(8):924-930 (in Chinese with English abstract).
- [21] Skalak DB. Prototype and feature selection by sampling and random mutation hill climbing algorithms. In: Cohen WW, Haym Hirsh, H, eds. Proc. of the Int'l Conf. on Machine Learning. Morgan Kaufmann Publishers, IEEE Press, 1994. 293-301.
- [22] Passerini A, Pontil M, Frasconi P. From margins to probabilities in multiclass learning problems. In: van Harmelen F, ed. Proc. of the 15th European Conf. on Artificial Intelligence, Lyon, France: IOS Press. 2002. 400-404.
- [23] Larsen J, Svarer C, Andersen LN, Hansen LK. Adaptive regularization in neural network modeling. In: Neural Networks: Tricks of the Trade, 1996.113-132.

附中文参考文献:

- [1] 陆汝钊.世纪之交的知识工程与知识科学.北京:清华大学出版社,2001.
- [2] 靳蕃.神经计算智能基础·原理·方法.成都:西南交通大学出版社,2000.
- [3] 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.
- [6] 丛爽.神经网络、模糊系统及其在运动控制中的应用.合肥:中国科学技术大学出版社,2001.
- [7] 陆汝钊.知识科学与计算科学.北京:清华大学出版社,2003.
- [9] 翟宜峰,李鸿雁,刘寒冰,苑希民.用遗传算法优化神经网络初始权重的方法.吉林大学学报(工学版),2003,33(2):45-50.
- [10] 殷峻暹,陈守煜,邱菊.基于遗传与BP混合算法神经网络预测模型及应用.大连理工大学学报,2002,42(5):594-598.
- [11] 郑志军,郑守淇.进化神经网络中的变异算子研究.软件学报,2002,13(4):726-731.<http://www.jos.org.cn/1000-9825/13/726.pdf>
- [12] 陶卿,曹进德,孙德敏,方廷健.基于约束区域神经网络的动态遗传算法.软件学报,2001,12(3):462-467.
- [13] 杨国军,崔平远,李琳琳.遗传算法在神经网络控制中的应用与实现.系统仿真学报,2001,13(5):567-570.
- [17] 郭爱克.计算神经科学.上海:上海科技出版社,2000.
- [20] 张军,刘克胜,王煦法.一种基于免疫调节和共生进化的神经网络优化设计方法.计算机研究与发展,2000,37(8):924-930.