

BP网络的SPDS算法的局部极小问题分析

张少仲¹, 李龙锁², 任世军², 蒋波¹, 白英¹, 张维石¹

(1. 大连海事大学信息科学技术学院, 116026 大连; 2. 哈尔滨工业大学计算机科学与技术学院, 150001 哈尔滨)

摘要: 针对BP网络的训练算法SPDS算法,研究了局部极小问题.利用基于单参数动态搜索算法的SPDS算法的变量逐一搜索的特点,证明了每次迭代的等价误差函数均为拟凸函数,进而极小点存在并可求出.将迭代必将收敛的初值集合定义为全局极小区域,针对局部极小问题给出L-SPDS算法,并证明了SPDS算法的全局极小区域沿坐标轴方向扩张的区域既是L-SPDS算法的全局极小区域,从而SPDS算法收敛于全局极小点的可能性大大增加了,算法的仿真试验也证明了这一点.

关键词: BP网络;SPDS算法;局部极小问题

中图分类号: TP39

文献标志码: A

文章编号: 0367-6234(2013)11-0125-04

Analysis of local minimum problem of SPDS algorithm in BP network

ZHANG Shaozhong¹, LI Longsuo², REN Shijun², JIANG Bo¹, BAI Ying¹, ZHANG Weishi¹

(1. Information Science and Technology College, Dalian Maritime University, 116026 Dalian, China;

2. School of Computer Science and Technology, Harbin Institute of Technology, 150001 Harbin, China)

Abstract: The local minimum problem of SPDS algorithm—the training algorithm of BP neural network, is studied. As one of SPDS algorithm features based on the single parameter dynamic searching algorithm is that the variables are searched one by one, this paper proves that the equivalence error function of each iteration is a quasi-convex function, and minimum points are presence and can be found out. The initial value set from which the iterative must be convergence is defined as the global minimum area, and according to the local minimum problems, L-SPDS algorithm is given. The global minimum area of the SPDS algorithm expanding along coordinate direction is the global minimum area of L-SPDS algorithm. The possibility that SPDS algorithm converges to the global minimum point greatly increases, which is proved by algorithm simulation test.

Key words: BP network; SPDS algorithm; Local minimum problem

BP算法的主要缺点之一是训练容易陷入局部极小点^[1].许多研究者采用随机优化技术,如模拟退火技术等^[2],来寻求全局最优解^[3-4].SPDS算法^[5-7]也存在局部极小值问题.但是,在解决一些实际问题的过程中,本文发现用SPDS算法收敛到全局极小点的可能性远大于BP算法.即使它收敛到局部极小点时,通过一些技术措施也容易使它收敛到全局极小点.

局部极小值的产生是与初始值的选取密切相关.

两个距离很近的初始点,可能最终一个将迭代到局部极小点,而另一个将迭代到全局极小点.本文将那些最终能迭代到全局极小点的初始点的集合定义为全局极小区域.本文首先从理论上提出L-SPDS算法,然后证明其全局极小区域比SPDS算法的全局极小区域大,由此不难看出,SPDS算法收敛到全局极小点的可能性大大增加了,这对于解决局部极小问题,也具有一定的贡献.

本文只研究了3层BP网络的误差函数为

$$E = \sum_p \sum_i (O_{p_i} - y_{p_i})^2.$$

式中: O_{p_i} 为输出层第*i*个节点对应于第*p*个训练样本的实际输出; y_{p_i} 为相应的期望输出值.

收稿日期: 2012-10-15.

基金项目: 国家自然科学基金资助项目(61173034).

作者简介: 张少仲(1965—),男,高级工程师.

通信作者: 张少仲, zhang_shaozhong2001@yahoo.com.cn.

$$E = \sum_p \sum_i \left(g \left(\sum_j w_{ij} \bar{b}_j + \theta_i \right) - y_{pi} \right)^2 = \sum_p \sum_i \left(g \left(\sum_j w_{ij} g \left(\sum_k w_{kj} b_k + \varphi_j \right) + \theta_i \right) - y_{pi} \right)^2. \tag{1}$$

式中： w_{ij} 为第 i 个输出层单元到第 j 个隐层单元的权值； w_{kj} 为第 k 个输入层单元到第 j 个隐层单元的权值； \bar{b}_j 为第 j 个隐层单元的输出； b_k 为第 k 个输入层单元的输出； θ_i 为第 i 个输出层单元阈值； φ_j 为第 j 个隐层单元阈值； g 为 Sigmoid 函数， $g' = \lambda g(1 - g)$ 。

神经网络的训练就是使 $\theta_i, w_{ij}, \varphi_j, w_{kj}$ 这 4 组变量取适当的值，使得误差函数 E 取得全局极小值。

基于单参数动态搜索算法的 SPDS 算法，就是将这 4 组变量逐一作为变量处理，其他值均用上次迭代后的储存值代替。这样 4 个误差函数分别记为 E_1, E_2, E_3 和 E_4 ，它们都是关于变量的连续可微函数。不难看出，每个误差函数均为以下形式：

$$E_i = \sum f_i^2(x).$$

式中， $f_i^2(x)$ 为一元函数， $i = 1, 2, 3, 4$ 。

1 全局极小区域

引理 1 $f_i^2(x)$ 均为 R^1 上的拟凸函数，其中 $i = 1, 2, 3, 4$ 。

证明 对于 E_1 ，即式(1)中，只有 θ_i 是变量（设为 x ），其余皆视为常量。故 $f_1(x) = g(\sum_j w_{ij} \bar{b}_j + x) - y_{p1}$ ，因为 g 为单调增连续函数，故 $f_1(x)$ 亦为单调增连续函数。

任给 $\alpha > 0, x_1, x_2 \in R^1$ ，若 $f_1^2(x_1) \leq \alpha$ ， $f_1^2(x_2) \leq \alpha$ ，则有

$$-\sqrt{\alpha} \leq f_1(x_1) \leq \sqrt{\alpha}, \\ -\sqrt{\alpha} \leq f_1(x_2) \leq \sqrt{\alpha}.$$

而 $\alpha > 0$ ，故存在 $y_1, y_2 \in R^1$ 使 $f_1(y_1) = -\sqrt{\alpha}$ ， $f_1(y_2) = \sqrt{\alpha}$ 且 $y_1 \leq x_1 \leq y_2, y_1 \leq x_2 \leq y_2$ 。

任意 $\lambda \in (0, 1)$ ，有 $y_1 \leq (1 - \lambda)x_1 + \lambda y_2 \leq y_2$ ，而 $f_1(x)$ 为单调增连续函数，故

$$-\sqrt{\alpha} = f_1(y_1) \leq f_1((1 - \lambda)x_1 + \lambda y_2) \leq f_1(y_2) = \sqrt{\alpha},$$

从而

$$f_1^2((1 - \lambda)x_1 + \lambda y_2) \leq \alpha.$$

所以，任意 $\alpha > 0, S(f_1^2, \alpha) = \{x \mid f_1^2(x) \leq \alpha\}$ 为凸集。故 $f_1^2(x)$ 为 R^1 上的拟凸函数。同理可证 $f_i^2(x)$ 为 R^1 上的拟凸函数，其中 $i = 2, 3, 4$ 。证毕。

定理 1 对于误差函数 E_i (其中 $i = 1, 2, 3, 4$)，若 $f_i^2(x)$ 在 R^1 上有极小值 x_i^* ，则 E_i 的极小值 x_i^0 满足：

$$\min x_i^* \leq x_i^0 \leq \max x_i^*.$$

证明 由引理知 $f_i^2(x)$ 为 R^1 上的拟凸函数，故存在区间 J_1 和 J_2 ， $\text{dom} f_i^2(x) = J_1 \cup J_2$ ，且 J_1 在 J_2 的左边， $f_i^2(x)$ 在 J_1 上不增，在 J_2 上不减^[8]，对于 $E_i = \sum f_i^2(x)$ ，当 $x < \min x_i^*$ 时， E_i 不增；当 $x > \max x_i^*$ 时， E_i 不减。因而 E_i 的极小值 x_i^0 满足：

$$\min x_i^* \leq x_i^0 \leq \max x_i^*.$$

定理得证。

由此可知， E_i (其中 $i = 1, 2, 3, 4$) 的极小点存在并可求出。

定义 1 在初始值可取区域中，迭代必然收敛到全局极小点的区域使本文成为全局极小区域。

2 L-SPDS 算法及全局极小区域扩张定理

对于局部极小问题，本文给出了 L-SPDS 算法：

1) 任给初值。

2) 按某种顺序，以下步骤逐个迭代上述 4 类参数：

a) 计算该参数的导数，得到误差函数沿该参数的下降方向。

b) 沿误差函数的下降方向进行搜索，直到导数零点停止。

3) 若迭代已落入局部极小点，则在初始点处求出误差函数沿任一参数的所有极小点，并分别以这些极小点为该参数的调整点，跳到步骤 2) 至步骤 3) 进行迭代；若误差函数 E 小于允许误差，则认为已收敛到全局极小点，迭代停止。

定理 2 SPDS 算法的全局极小区域沿坐标轴方向延长形成新的区域，该新的区域是 L-SPDS 算法的全局极小区域。

证明 设 A 点为 SPDS 算法的全局极小区域中任一固定点， B 点为 A 点沿坐标轴方向延长形成一条直线上的任意点，只要证明以 B 点为初始点，L-SPDS 算法将收敛于全局极小点即可。

设 C 点是 SPDS 算法以 A 点为初始点时，目标函数沿该坐标轴方向迭代的下一个极小点。由 C 点出发继续迭代将收敛到全局极小点。以 B 点为初始点时，沿该坐标轴方向可求出每个极小点（其中包括 C 点）。若分别以这些极小点为调整的下一点进行迭代的话，由于以 C 点为迭代的下一

点进行的迭代将收敛于全局极小点, 故这样的迭代将至少有一次收敛于全局极小点. 而这正是 L-SPDS 算法的迭代过程, 定理得证.

定理 2 中, 由 C 点出发继续进行迭代将收敛于全局极小点. 具有这种性质的 C 点往往较多. 这样的迭代, 计算量有相当的增加, 但全局极小区域却扩大了. 以误差函数是二元函数的情况为例, 图 1 为二元函数的取值范围, 设区域 Ω_1 和区域 Ω_2 均为 SPDS 算法的全局极小区域, 则阴影以外的部分 (包括区域 Ω_1 和区域 Ω_2) 都是 L-SPDS 算法的全局极小区域. 在实际计算中, L-SPDS 算法中的求出某参数的所有极小点有时可改为求得某个极小点或者用随机优化技术来替代. 此时 L-SPDS 算法即转化为 SPDS 算法. 在一维情况下, 使用随机优化技术搜索到定理 2 中的 C 点相对比较容易实现. 因此, SPDS 算法收敛到全局极小点的可能性比较大. 而 BP 算法却无此特性, 这是对 BP 网络训练算法的贡献.

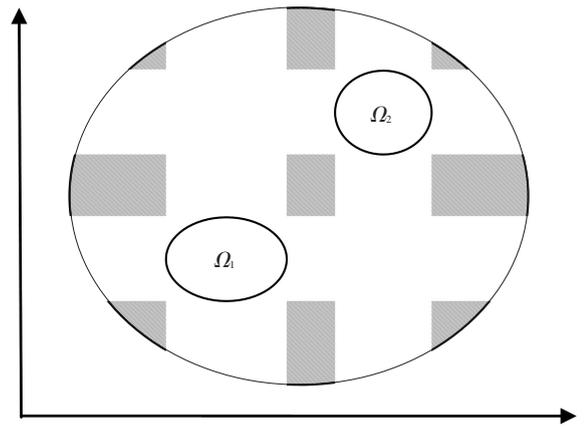


图 1 全局极小区域扩张示意图

3 仿真算例

例 1 以 Xor 问题为例. 网络结构为 2-2-1, 初值选取如表 1 所示. 在迭代到第 19 步, 调整第 1 个输入层单元与第 1 个隐层单元的权值时, 误差函数与该权值的关系如图 2 所示, 图 3 为误差函数关于该权值的导数与该权值的关系.

表 1 Xor 问题的初始值

单元	权值		阈值
	输入层第 1 单元	输入层第 2 单元	
隐层第 1 单元	9.511 414 e-003	2.705 688e-003	3.747 559e-003
隐层第 2 单元	-8.410 645 e-004	9.068 298e-003	-9.447 937e-003

单元	权值		阈值
	隐层第 1 单元	隐层第 2 单元	
输出层单元	4.474 182 e-003	1.269 531e-003	9.968 517e-001

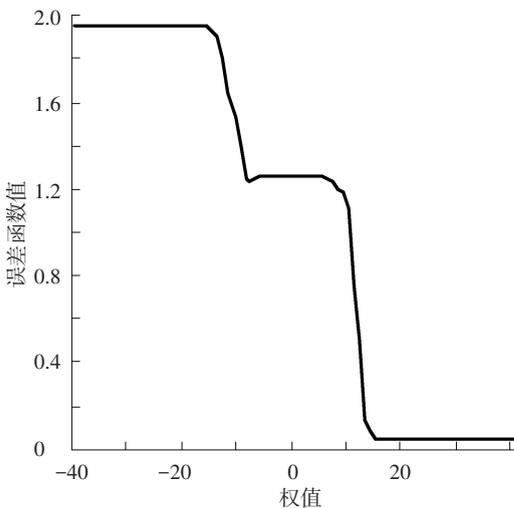


图 2 误差函数与该权值的关系

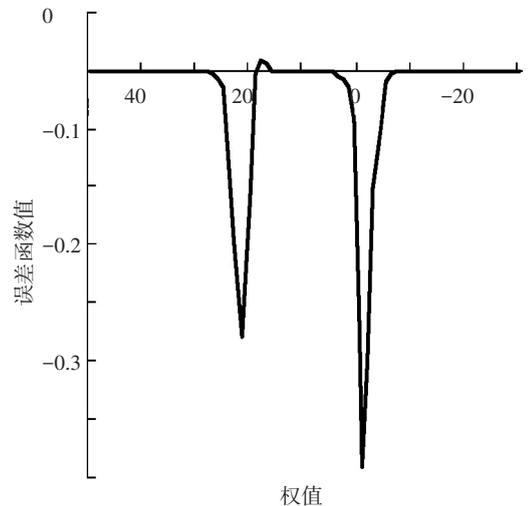


图 3 误差函数关于该权值的导数与该权值的关系

从图 2 中可知, -18.9 和 6.39 均为误差函数沿该权值的极小点. 在图 3 中也能看到同样的结果. 分别选择这两个极小点作为该权值的下一步调整值, 继续进行迭代, 得到结果如表 2、

3 所示.

从表 2 中可看出, 选择极小点 6.39 作为该权值的下一步调整值, 迭代结果很快收敛; 而从表 3 中可知, 选择极小点 -18.9 作为该权值的下一步

调整值,迭代结果不收敛,可以认为已陷入局部极小.

表 2 以 6.39 为调整值的迭代结果

极小点值	极小点处函数值
6.39	$4.72e-2$
迭代次数	均方误差
19	$4.47e-2$
40	$1.24e-7$
50	$1.20e-10$

表 3 以 -18.9 为调整值的迭代结果

极小点值	极小点处函数值
-18.9	$1.23e-0$
迭代次数	均方误差
19	$1.15e-0$
40	$5.02e-1$
100	$5.01e-1$
500	$5.00e-1$
5 000	$5.00e-1$
10 000	$5.00e-1$

即-18.9 对应的参数点不属于 SPDS 算法的全局极小区域,6.39 对应的参数点属于其全局极小区域.在进行到此处的迭代时,分别选取 6.39 和-18.9 作为该权值的下一个调整值进行迭代,将又一次收敛到全局极小点,即-18.9 对应的参数点属于 L-SPDS 算法的全局极小区域,亦能收敛到全局极小点.

由此不难看出,SPDS 算法收敛到全局极小点的可能性大大增加了,而 BP 算法是将所有参数作为一个整体进行统一搜索的,没有此性质.

4 结 论

1)对于 BP 网络的 SPDS 算法,证明了每次迭代的等价误差函数均为拟凸函数,进而极小点存在并可求出.

2)将迭代必将收敛的初值集合定义为全局极小区域,针对局部极小问题给出 L-SPDS 算法,并证明了 SPDS 算法的全局极小区域沿坐标轴方向扩张的区域既是 L-SPDS 算法的全局极小区域.从而 SPDS 算法收敛于全局极小点的可能性大大增加了.

参考文献

[1] 张铃,张钊. 神经网络中 BP 算法的分析[J]. 模式识别与人工智能,1994,7 (3): 191-195.

[2] 康立山,谢云,尤矢勇,等. 非数值并行算法:模拟退火算法[M]. 第 1 册. 北京:科学出版社,1994.

[3] LO J T H, GUI Yichuan, PENG Yun. Overcoming the local-minimum problem in training multilayer perceptrons with the NRAE training method [C]// Proceedings of the 9th international conference on Advances in Neural Networks-Volume Part I. Berlin, Heidelberg: Springer-Verleg, 2012: 440-447.

[4] CHEUNG C C, NG S C, LUI K A. Multi-phase fast learning algorithms for solving the local minimum problem in feed-forward neural networks [C]// Proceedings of the 9th international conference on Advances in Neural Networks-Volume Part I. Berlin, Heidelberg: Springer-Verleg, 2012: 580-589.

[5] WANG Xuefeng, FENG Yingjun, ZHAO Xin. A single parameter dynamic searching algorithm for multi-layer neural networks (Part I) [J]. International Journal of Innovative Computing, Information and Control, 2008, 4 (9): 2221-2233.

[6] WANG Xuefeng, ZHAO Xin, FENG Yingjun. A single parameter dynamic searching algorithm for multi-layer neural networks (Part II) [J]. International Journal of Innovative Computing, Information and Control, 2008, 4 (11): 2941-2954.

[7] 王雪峰,冯英俊. 多层神经网络的一种新的学习算法 [J]. 哈尔滨工业大学学报,1997, 29 (2): 23-25.

[8] 刘光中. 凸分析与极值问题[M]. 北京:高等教育出版社,1991.

[9] 张少仲,冯英俊,王雪峰,等. 基于单参数坐标轮换法的多层前馈型神经网络的一种新型算法[J]. 电子学报,1999, 27 (2): 115-116.

[10] 张少仲,冯英俊. 单参数动态搜索算法的泛化与容错能力[J]. 哈尔滨工业大学学报,2000, 32 (6): 1-4.

[11] 冯英俊,张少仲,王雪峰,等. BP 网络瘫痪的原因分析[J]. 哈尔滨工业大学学报,2000, 32 (5): 81-83.

[12] 张少仲,冯英俊. BP 网络的 SPDS 训练算法的鲁棒性[J]. 哈尔滨工业大学学报,2000, 32 (2): 60-61.

(编辑 张 红)