

BP 网络不平衡训练样本集的有效学习算法¹

江铭虎 林碧琴 袁保宗 许晓斌

(北方交通大学信息科学研究所 北京 100044)

摘要 分析了 BP 网络标准反传学习算法对不平衡样本集训练速度慢的原因, 研究了如何改进其学习算法来加速训练速度, 并通过实验对上述理论进行验证。

关键词 神经网络, 不平衡训练集

中图分类号 TN-052

1 引言

BP 网络反传算法是减少实际输出与目标输出的欧氏距离^[1], 但由于实际工作中经常遇到不平衡训练样本集^[2], 它的某些参数其作用抑制着输出误差的变化, 这样反过来使权变化减少, 训练时间加长。为解决这一问题, 下面分析训练过程中由不平衡样本集而引起收敛速度减慢的原因, 并研究如何改进其学习算法来加速训练速率。

2 不平衡训练样本集对学习速率的影响及改进方案

设双隐层 BP 网络第 k 类第 j 样本的输入为 $U^{(j,k)} = (u_1^{(j,k)}, \dots, u_{I+1}^{(j,k)})$, 隐层 1 第 l 节点的权值为 $H_l = (h_{l,1}, \dots, h_{l,I+1})$, 该层输出为 $X^{(j,k)} = (x_1^{(j,k)}, \dots, x_{M+1}^{(j,k)})$, 隐层 2 第 r 节点的权值为 $W_r = (W_{r,1}, \dots, W_{r,M+1})$, 该层输出为 $Y^{(j,k)} = (y_1^{(j,k)}, \dots, y_{N+1}^{(j,k)})$, 输出层第 s 节点的权值为 $V_s = (V_{s,1}, \dots, V_{s,N+1})$, 目标输出为 $t_s^{(j,k)} = \begin{cases} 1, & s = k, \\ 0, & s \neq k, \end{cases}$ 该层输出为 $Z^{(j,k)} = (z_1^{(j,k)}, \dots, z_K^{(j,k)})$, 激活函数为 sigmoid() 函数。

BP 算法权系数调整是按 $W_z(n+1) = W_z(n) - \lambda \nabla E(W_z(n))$ 进行的, 其中 $W_z = (V, W, H)$, ∇E 为误差梯度, n 为叠代次数。每层由误差反传引起的权值变化量如下^[1]:

隐层 2 第 r 节点至输出层第 s 节点的权变化为

$$\Delta v_{s,r}^{(j,k)} = \lambda \delta_s^{(0)} y_r^{(j,k)} = \lambda (t_s^{(j,k)} - z_s^{(j,k)}) z_s^{(j,k)} (1 - z_s^{(j,k)}) y_r^{(j,k)}. \quad (1)$$

隐层 1 第 l 节点至隐层 2 第 r 节点的权变化为

$$\Delta w_{r,l}^{(j,k)} = \lambda \delta_r^{(2)} x_l^{(j,k)} = \lambda y_r^{(j,k)} (1 - y_r^{(j,k)}) \sum_{m=1}^K \delta_m^{(0)} v_{m,r} x_l^{(j,k)}. \quad (2)$$

同理可推得输入层 q 节点至隐层 1 第 l 节点的权变化为

$$\Delta h_{l,q}^{(j,k)} = \lambda \delta_l^{(1)} u_q^{(j,k)} = \lambda [x_l^{(j,k)} (1 - x_l^{(j,k)}) \sum_{n=1}^N \delta_n^{(2)} w_{n,l}] u_q^{(j,k)}. \quad (3)$$

假设在训练初期对两类分类问题。

¹ 1997-10-22 收到, 1998-06-26 定稿
国家自然科学基金资助课题 (NO.69672002)

定理 1 $\Delta v_{s,r}^{(j,k)} \begin{cases} > 0, & \text{当 } s = k \text{ 时,} \\ < 0, & \text{当 } s \neq k \text{ 时.} \end{cases}$

证明 因为输出层 s 节点的期望值为 $t_s^{(j,k)} = \begin{cases} 1, & \text{当 } s = k \text{ 时,} \\ 0, & \text{当 } s \neq k \text{ 时.} \end{cases}$ 由激活函数为 sigmoid()

函数知 $0 < y_s^{(j,k)} < 1, 0 < z_s^{(j,k)} < 1$, 再由 (1) 式及 $\lambda > 0$ 便得结论. 这说明就两类问题对于某一类所有样本的输出层权变化有同样的符号变化, 而另一类具有相反的符号变化.

定理 2 由定理 1 及前面的假设, 有 $\nabla E_1(W_z) \cdot \nabla E_2(W_z) < 0$.

证明 对输出节点 s , $\nabla E_1(W_z) \cdot \nabla E_2(W_z)$ 的点积为

$$\nabla E_1(W_z) \cdot \nabla E_2(W_z) = \sum_{r=1}^{N+1} \Delta v_{s,r}^{(1)} \Delta v_{s,r}^{(2)} + \sum_{r=1}^{N+1} \sum_{l=1}^{M+1} \Delta w_{r,l}^{(1)} \Delta w_{r,l}^{(2)} + \sum_{l=1}^{M+1} \sum_{q=1}^{I+1} \Delta h_{l,q}^{(1)} \Delta h_{l,q}^{(2)}. \quad (4)$$

从定理 1 及 (1), (2), (3) 式知, 点积的每一项 $(\Delta v_{s,r}^{(1)}, \Delta v_{s,r}^{(2)})$, $(\Delta w_{r,l}^{(1)}, \Delta w_{r,l}^{(2)})$ 和 $(\Delta h_{l,q}^{(1)}, \Delta h_{l,q}^{(2)})$ 中, 其符号仅取决于输出层的 $t - z$, 可得出某一因子为正, 另一因子必为负, 因此点积总为负, 即 $\nabla E_1(W_z) \cdot \nabla E_2(W_z) < 0$, 说明 $\nabla E_1(W_z)$ 与 $\nabla E_2(W_z)$ 间的夹角总是大于 90° .

由 $E_k(W_z) = \sum_{j=1}^{n_k} (t^{(j,k)} - z^{(j,k)})^2 / 2$, 其中 $j = 1, 2, \dots, n_k; k = 1, 2$. 知反传梯度误差的幅度是与训练样本的数量成正比的, 占优势的样本集误差幅度变化大, 占劣势的样本集误差幅度变化小.

设第 k 类第 j 个样本梯度矢量的变化为

$$\nabla E^{(j,k)}(W_z) = (\Delta V^{(j,k)}, \Delta W^{(j,k)}, \Delta H^{(j,k)}) \text{ 和 } \nabla E_k(W_z) = \sum_j \nabla E^{(j,k)}(W_z), k = 1, 2.$$

通过实验观察发现, 在训练初期占优势的样本类反传误差减少很快, 但占劣势的样本类反传误差反而增加, 之后占劣势的误差减少很慢, 从而导致慢的收敛速率. 见图 1 所示, 设第二类样本远远多于第一类样本, 标准的误差学习算法中矢量 AD 是减少 $E(W_z)$ 的最佳方向, 而 AD 可分解为 AE (对应于 $E_1(W_z)$) 和 AF (对应于 $E_2(W_z)$), 但矢量 AE 是使 $E_1(W_z)$ 增加的方向, 结果矢量 AD 方向权值的变化使 $E_2(W_z)$ 迅速减少, 而使 $E_1(W_z)$ 迅速增加. 反传误差变化的幅度是与矢量 AF 和 AE 的长度成正比的. 在训练的初始阶段, 训练误差较大, 它对应于 $t = 1, z \approx 0$ 或 $t = 0, z \approx 1.0$, 此时有 $|t - z| \approx 1$ 和 $(t - z)z(1 - z) \approx 0$. 由 (1) 式知反传梯度误差改变的幅度是与 $(t - z)z(1 - z)$ 成正比的, 此时权值改变的幅度很小, 需大量的迭代次数, 训练很慢. 从图 1 知, 经初次迭代后 $E_1(W_z)$ 变得很大, 由上面的分析知 $E_1(W_z)$ 的权值改变量很小, 收效速率很慢; 而 $E_2(W_z)$ 原先很大, 其权值变化很小, 结果标准误差反传算法对于不平衡训练样本集训练速度很慢.

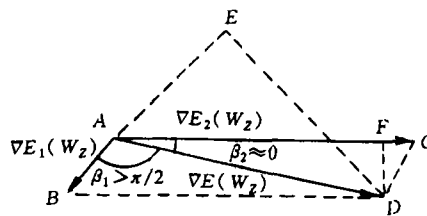


图 1 梯度矢量 $\nabla E(W_z), \nabla E_1(W_z), \nabla E_2(W_z)$ 的关系

从上面的分析知, 就不平衡样本的两类问题, $-\nabla E(W_z)$ 并不总是指向同时使两类误差减少的最佳方向, 因此需采用改进的梯度反传学习算法来纠正。

方法 1 按 $W(n+1) = W(n) - \lambda\alpha(n)$ 进行叠代, 并使 $\alpha(n)$ 满足^[2]: $-\alpha(n) \cdot \nabla E_k(W(n)) < 0$, 其中 $k = 1, 2$ 。迭代次数为 n 。

除非 $\nabla E_1(W_z)$ 与 $\nabla E_2(W_z)$ 的夹角为 180° , 总能找到使 $E_1(W_z)$ 和 $E_2(W_z)$ 同时下降的梯度矢量。这一梯度矢量(如图 2 中 AD' 即为 $\alpha(n)$) 与 AB 和 AC 之间的夹角均小于 90° 。满足图 2 条件的 α 矢量可选为

$$-\alpha(n) \cdot \nabla E_1(W_z(n)) / \|\nabla E_1(W_z(n))\| = -\alpha(n) \cdot \nabla E_2(W_z(n)) / \|\nabla E_2(W_z(n))\|, \quad (5)$$

$$\|\alpha(n)\| = (\|\nabla E_1(W_z(n))\| + \|\nabla E_2(W_z(n))\|) / 2. \quad (6)$$

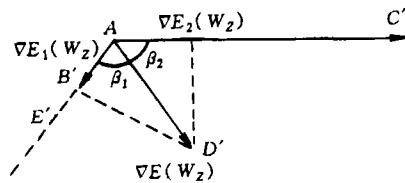


图 2 梯度矢量改进有效算法

方法 2 对少数样本集的进行样本选择性重复使用, 使两类样本集的数量大致相同。

这两种算法避免了标准反传算法的弊端, 每次迭代能快速地使两类误差按同时减小的方向变化。同理可推广到多类问题的分类。

3 对比实验

(1) 实验是以语音信号的前 5 帧归一 12 阶 LPC 倒谱系数作为输入特征, 语音样本是对 b 、 d 、 g 三音呼 20 遍共 60 个样本进行训练, 区分两类样本 b 和 $(d、g)$ 的 BP 网络为 $(61, 21, 7, 1)$, 其中 61, 21, 7, 1 分别是输入层, 隐层 1, 2 和输出层的节点数, 训练次数和识别结果如表 1 所示, 可见改进算法的训练时间仅为标准算法的 $1/8$, 而识别率基本相当。

表 1 不平衡训练样本集标准算法与改进算法的比较

	标准算法	改进算法 1	改进算法 2
训练时间 (s)	2159	268	243
集外识别率	96.4%	96.7%	96.5%

(2) 实验是语音信号的 16 阶 Mel 谱系数, 取前 15 帧进行峰值归一作为输入特征。语音样本是对 $\{b、d、g、p、t、k\}$ 呼 10 遍共 60 个样本分两组辅音集进行训练, 网络是采用 Waibel 提出的时延神经网络^[3,4], 它是按层误差反传。首先对 b 和 $(d、g)$, p 和 $(t、k)$ 及 b 和 $(d、g、p、t、k)$ 语音不平衡样本集分别进行对比训练, 实验结果如表 2 所示。

表 2 时延神经网络不平衡样本训练集标准学习算法与改进学习算法的比较

TDNN 网络	$\{b\}, \{d, g\}$			$\{p\}, \{t, k\}$			$\{b\}, \{d, g, p, t, k\}$		
	标准算法	改进算法 1	改进算法 2	标准算法	改进算法 1	改进算法 2	标准算法	改进算法 1	改进算法 2
训练时间 (min)	86	17	16	88	18	17	279	18	17
集外识别率	99.2%	99.1%	99.2%	99.3%	99.4%	99.2%	98.3%	98.2%	98.3%

4 结 论

BP 网络反传梯度误差的幅度是与训练样本的数量成正比的, 它的计算受占优势样本集的制约。对不平衡训练样本集在训练的初始阶段, 训练误差较大, 经初次迭代后使占少数样本集类的反传误差变得很大, 其权值改变量很小, 收效速率很慢; 而占多数样本集类的反传误差原先很大, 其权值变化很小, 结果标准误差反传算法对于不平衡训练样本集收敛速度很慢。我们提出改进算法每次迭代能快速地按两类误差同时减小的方向变化, 它能有效地提高学习速率。

参 考 文 献

- [1] Anand R, *et al.* Efficient classification for multiclass problem using modular neural networks. *IEEE Trans. on Neural Networks*, 1995, NN-6(1): 117-124.
- [2] Anand R, *et al.* An improved algorithm for neural networks classification of imbalanced training sets. *IEEE Trans. on Neural Networks*, 1993, NN-4(6): 962-969.
- [3] Waibel A, *et al.* Modularity and scaling in large phonemic neural networks. *IEEE Trans. on ASSP*, 1989, ASSP-37(12): 1888-1898.
- [4] Waibel A, *et al.* Phoneme recognition using time-delay neural networks. *IEEE Trans. on ASSP*, 1989, ASSP-37(3): 328-339.

AN EFFICIENT LEARNING ALGORITHM OF NONBALANCED TRAINING SET FOR BP NETWORK

Jiang Minghu Lin Biqin Yuan Baozong Xu Xiaobin

(*Institute of Information Science, Northern Jiaotong University, Beijing 100044*)

Abstract This paper analyzes the reason for low convergence rate of BP network standard backpropagation algorithm to nonblanced training set, researches how to improve the learning algorithm in order to increase training speed, and through two experiments tests and verifies above theory.

Key words Neural network, Nonblanced training sets

江铭虎: 男, 1962 年生, 博士, 从事神经网络语音处理的研究.

林碧琴: 女, 1942 年生, 副教授, 从事语音处理的研究.

袁保宗: 男, 1932 年生, 教授, 博士生导师, 国务院学位委员, IEE Fellow, IEEE 计算机北京分会主席, 从事语音、图象信号处理领域的研究.

许晓斌: 男, 1969 年生, 博士生, 从事语音处理的研究.