

文章编号: 1007-4627(2014) 03-0401-06

基于共轭梯度优化算法的 BP 神经网络 在高能粒子鉴别等领域中的应用

王树旺¹, 路永钢¹, 陈旭荣^{2, 3}

(1. 兰州大学信息科学与工程学院, 兰州 730000;

2. 中国科学院近代物理研究所, 兰州 730000;

3. 兰州强子物理和 CSR 物理研究中心, 兰州 730000)

摘要: 人工神经网络方法已被引入高能物理实验领域并被广泛地应用于夸克胶子喷注的鉴别、电子强子分辨、顶夸克和 Higgs 粒子的寻找等等。本文采用了一种改良的共轭梯度优化算法并应用于高能物理实验中粒子的鉴别。在该应用中, 此算法既能实现每步迭代时在搜索方向上获得最优步长, 又能避免目标函数陷入局部收敛点, 从而使目标函数快速收敛, 提高了算法的有效性。分析结果表明, 我们改进后的 BP 算法显著地提高了粒子物理数据分析中的粒子鉴别能力。

关键词: BP 神经网络; 共轭梯度法; 步长优化; 粒子鉴别

中图分类号: TP183 **文献标志码:** A **DOI:** 10.11804/NuclPhysRev.31.03.401

1 引言

在大型粒子物理实验数据分析中, 粒子鉴别问题是需要解决的最基本最重要的问题之一^[1]。粒子鉴别问题实质是统计分析中的模式识别问题。近年来, 人们发展了很多用于解决此类问题的方法, 比如贝叶斯方法^[2]、决策树法^[3]、k 近邻法^[4]、线性判别方法等。其中应用最为广泛的是 Fisher 线性判别方法^[5]和神经网络方法^[6]。上述方法在预测能力、可解释性、稳定性、训练以及分类所消耗的时间上都有各自的优势, 在实际应用中我们要根据分析的实际需要来选择合适的分类器。

人工神经网络 (artificial neural networks, NN) 由于自身的优点, 已经在信息处理、模式识别、智能控制及系统建模等领域得到越来越广泛的应用。早在 1988 年, 人工神经网络方法就被引入高能物理实验领域^[7-8]并广泛地应用于夸克胶子喷注的鉴别^[9-11]、电子强子分辨、顶夸克和 Higgs 粒子的寻找等等。目前, 在模式识别中应用较成熟的神经网络模型是误差反向传递 (Back-Propagation) 算法的前馈

网络 (简称 BP 网络)。BP 网络可以以任意精度逼近任意的连续函数, 所以广泛应用于非线性建模、函数逼近、模式分类等方面。但是, 传统的 BP 算法^[12]在学习过程中, 学习步长的选取^[13-14]一般是人为确定的, 并且是固定不变的, 这是一个很大的缺点。究其原因总是误差 E 是一个十分复杂的非线性函数, 很难通过解析的方法得到最优步长。如果学习步长选得太大, 会使学习优化产生振荡甚至发散, 如果学习步长选得太小, 不仅会增加迭代次数, 而且会陷入局部极小值, 最终搜索不到最优点。所以如何确定一个最优步长, 是值得深入研究的重要问题。另外, 传统的 BP 算法采用的是负梯度算法, 该算法存在收敛速率低, 训练时间长等缺点, 严重影响了 BP 网络的应用。因此, 我们尝试改用学习效率较高的共轭梯度法, 对于推广 BP 网络的应用, 具有十分重要的意义。

2 共轭梯度优化算法及步长优化

2.1 共轭梯度法

共轭梯度法是一种改进搜索方向的方法, 它是

收稿日期: 2013-10-11; 修改日期: 2014-01-08

基金项目: 国家自然科学基金(11005045, 1117522, 61272213); 中国科学院百人计划(Y101020BR0)

作者简介: 王树旺(1981-), 男, 汉族, 甘肃白银人, 硕士研究生, 从事神经网络研究; E-mail: wangsw.lzu@163.com

通信作者: 陈旭荣, E-mail: xchen@impcas.ac.cn.

<http://www.npr.ac.cn>

把前一步的梯度乘以适当的系数，加到当前的梯度上，得到新的搜索方向。因此，可以说共轭梯度法是把过去的梯度信息和当前的梯度信息综合利用，来构造更优的搜索方向，新的搜索方向可以用公式表示为 $d_{i+1} = -\nabla f(x_{i+1}) + \beta_i d_i$ ，式中 β_i 为改变搜索方向的方向因子，可由下式确定：

$$\beta_i = \frac{\|\nabla f(x_{i+1})\|^2}{\|\nabla f(x_i)\|^2} \tag{1}$$

其中： d_i 为第 i 步搜索方向， d_{i+1} 为第 $i+1$ 步的搜索方向，且 d_i 和 d_{i+1} 必须互为共轭方向。

共轭梯度法的算法步骤如下：

- (1) 确定初始点 x_0 ，给定终止误差 $\varepsilon > 0$ ；
- (2) 求 $\nabla f(x_0)$ ，使初始搜索方向 $d_0 = -\nabla f(x_0)$ ，令 $\varepsilon > 0$ ；
- (3) 从 x_i 点出发，沿 d_i 方向进行搜索求出 $x_{i+1} = x_i + \lambda_i d_i$ ，其中步长 λ_i 可使 $f(x_i + \lambda_i d_i) = \min f(x_i + \lambda d_i)$ ，且 $\lambda > 0$ ；
- (4) 求出 $d_{i+1} = -\nabla f(x_{i+1}) + \beta_i d_i$ ；
- (5) 若 $\|\nabla f(x_{i+1})\| < \varepsilon$ ，则迭代停止，否则令 $i = i + 1$ ，转到第 (3) 步进行计算。

共轭梯度算法是建立在二次模型上，具有二次终止性，是一种有效的算法，算法原理简单，易于编程，既克服了负梯度算法易陷入局部收敛点的缺点，又避

免了牛顿法计算量大的缺点。但是，该算法必须以负梯度方向为初始搜索方向，以 N 次迭代为一轮。缺点就是该算法在学习过程中，步长选择不当严重制约了神经网络计算的有效性^[15-16]，所以本文采用了一种新的方法，通过二次插值法来确定最佳步长，从而对共轭梯度法做进一步的优化。

2.2 步长优化

2.2.1 二次插值法

二次插值法就是用一个二次函数来逼近 $f(x)$ 的方法。其基本思想是在搜索区间中不断用低次(一般不超过三次)插值多项式来近似为目标函数，并逐步用插值多项式的极小点来逼近函数的极小点。设二次函数 $f(x)$ 经过 $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ 三点，且满足 $x_1 < x_2 < x_3$ ， $y_1 > y_2$ ， $y_2 < y_3$ ，将该三点用一条二次曲线来拟合，如图 1 所示。

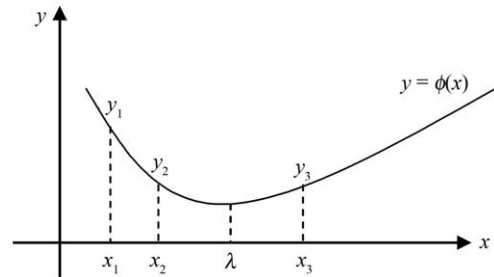


图 1 二次拟合曲线

推导出二次函数 $y = \phi(x)$ ，则

$$\phi(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)}y_1 + \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)}y_2 + \frac{(x-x_2)(x-x_1)}{(x_3-x_2)(x_3-x_1)}y_3 \tag{2}$$

求 $\phi(x)$ 的极小值，令 $d\phi(x)/dx = 0$ ，得

$$\lambda = \frac{1}{2} \frac{(x_3^2 - x_2^2)y_1 + (x_1^2 - x_3^2)y_2 + (x_2^2 - x_1^2)y_3}{(x_3 - x_2)y_1 + (x_1 - x_3)y_2 + (x_2 - x_1)y_3} \tag{3}$$

则 λ (如图 1 所示) 可近似看做函数 $f(x)$ 在区间 (x_1, x_3) 内的最小值。

2.2.2 步长优化

为了更加简洁、明确地说明BP算法中利用二次插值法确定最佳步长的计算方法，我们可采用流程图来表示计算方法的步骤，如图 2 所示，其中， E 为总体平均误差， t_k 为期望输出； $z_k = z_k(w_k, \lambda_k)$ 为实际输出； λ 为我们所求的最佳步长，由于 Δw 始终为负梯度方向，所以该算法中的循环总能在有限步内终

止，并找到最佳步长 λ 。

2.3 共轭梯度优化算法

我们将最佳步长计算的方法和共轭梯度法相结合，设计出一种新的共轭梯度优化算法，该算法通过二次插值的方法在学习过程中确定最佳学习步长，克服了共轭梯度法靠人为经验给定步长的缺陷。但是，

新的算法也必须以负梯度方向为初始搜索方向, 以 N 次迭代为一轮。算法具体步骤如下:

- (1) 选取初始点 x_0 , 并设可接受的误差 $\varepsilon > 0$;
- (2) 计算 $\nabla f(x_0)$, 使初始搜索方向 $d_0 = -\nabla f(x_0)$, 令 $i = 0$;

(3) 若 $\|\nabla f(x_i)\| < \varepsilon$, 则迭代停止, 否则沿 d_i 进行搜索, 求出 $x_{i+1} = x_i + \lambda_i d_i$, 式中 λ_i 的由 2.2.2 节描述的最佳步长法所确定;

(4) 求出 $d_{i+1} = -\nabla f(x_{i+1} + \beta_i d_i)$, 令 $i = i + 1$, 转到第 (3) 步进行计算。

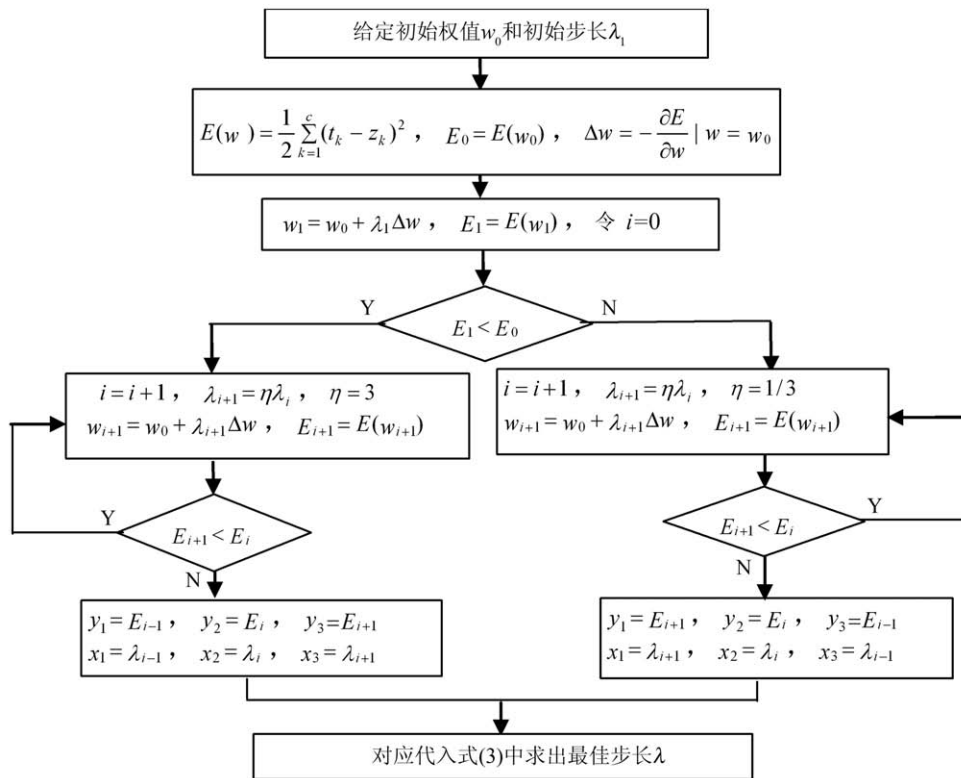


图 2 步长优化流程图

3 实验数据的判别分析

3.1 神经网络设置

用神经网络辨别胶子和夸克喷注是高能物理中很典型的例子。本实验样本^[17]是由蒙特卡罗事件产生器产生的质心能量为 91.2 GeV 的正负电子碰撞事件。在这一碰撞能量下正负电子湮灭产生一个高能虚光子或 Z 玻色子, 然后这些中间粒子衰变成两个正反夸克, 当正反夸克能量较高时有可能辐射出一个具有较高横能量的胶子。夸克和胶子在末态碎裂为强子, 从而在探测器中形成末态三喷注事件。数据分析中, 我们需要解决的问题是确定末态强子喷注来源于夸克还是胶子。在实验中无法直接记录夸克和胶子(部分子)如何强子化到强子, 但在蒙特卡罗模拟中可以将部分子层次和强子层次对比, 通过部分子信息得到强子喷注的

类别, 从而得到训练神经网络的夸克喷注和胶子喷注样本。本实验数据我们记录了喷注的七个特征参量: 分别为喷注多重数 n 、喷注内粒子平均横动量 p_t 、喷注所对的夹角 θ 、喷注能量 ε 、喷注内粒子的总横动量 pt_{total} 、喷注内粒子相对喷注轴的平均夹角 θ_{avg} 以及喷注内粒子相对喷注轴的夹角之和 θ_{avsum} 。在夸克胶子喷注的识别中, 选择反映两者不同性质的特征物理量至关重要。

我们选用只有一个隐含层的 BP 网络来鉴别夸克胶子喷注, BP 网络的的激活函数为 $f(x) = 1/(1 + e^{-2.5x})$ 。输入神经元个数由夸克胶子喷注的特征物理量个数决定, 在这里是 3, 对于神经网络的隐含层和输出层神经元数, 在理论上没有严格规定, 通常是固定不变的。这里为了研究神经元数目对网络识别喷注灵敏度和特异性的影响, 我们将隐含层神经

元数设为 6 或 2，输出层神经元数设为 2，分别对不同结构的神经网络进行训练并计算其判别效率。这里需要定义一个训练收精度 ϵ ，当 $E < \epsilon$ 时，停止训练并保存最新的网络权值和阈值，然后用这个训练好的网络来鉴别未知的夸克和胶子喷注。在训练过程中，网络平均误差 E 随着训练次数的收敛趋势如图 3 所示。

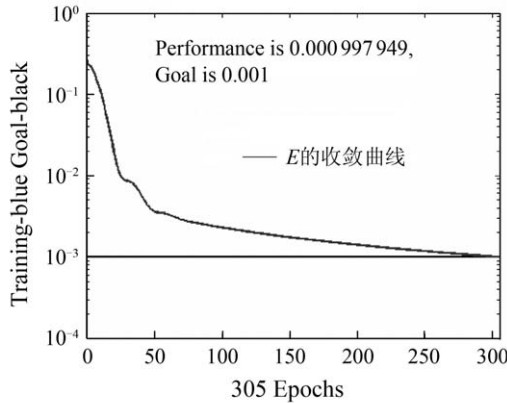


图 3 在训练过程中网络平均误差 E 的收敛图

假设神经网络的两个输出分别为 o_1 和 o_2 ，由于 BP 网络采用的激活函数是 Sigmoid 型的，这使得网络的输出值在 $(0, 1)$ 之间。若输入网络的是胶子喷注，则 o_1 接近于 1 且 o_2 接近于 0；若输入网络的是夸克喷注，则 o_1 接近于 0 且 o_2 接近于 1，于是我们定义一个收敛度 σ ：

若 $o_1 > 1 - \sigma$ 且 $o_2 < \sigma$ ，神经网络判定该喷注为胶子喷注；

若 $o_1 < \sigma$ 且 $o_2 > 1 - \sigma$ ，神经网络判定该喷注为夸克喷注；

若 $1 - \sigma > o_1 > \sigma$ 或 $1 - \sigma > o_2 > \sigma$ ，神经网络无法判定该喷注。

在夸克和胶子的判别问题中，通常选用判别效率、灵敏度和特异性来评价分类的好坏，假设有 n 个胶子喷注和 m 个夸克喷注组成的测试样本，经过神经网络判别后，有 n_g 个喷注被判定为胶子， m_q 个喷注被判定为夸克， n_{fail} 个无法判别。其中，在被判定为胶子的 n_g 个喷注中有 n_{cg} 个是正确判定为胶子，在判定为夸克的 m_q 个喷注中有 m_{cq} 个是正确判定的夸克，则定义式如下：

$$\text{判定效率: } t_{\text{eff}} = \frac{n_g + m_q}{n_g + m_q + n_{\text{fail}}},$$

$$\text{胶子特异性: } j_{\text{sp}} = \frac{m_{\text{cq}}}{m_{\text{cq}} + n_g - n_{\text{cq}}},$$

$$\text{夸克特异性: } k_{\text{sp}} = \frac{n_{\text{cg}}}{n_{\text{cg}} + m_q - m_{\text{cq}}},$$

$$\text{胶子灵敏度: } j_{\text{se}} = \frac{n_{\text{cg}}}{n},$$

$$\text{夸克灵敏度: } k_{\text{se}} = \frac{m_{\text{cq}}}{m}.$$

判定效率用来衡量神经网络判别粒子喷注的能力，是指测试样本中能被判别的粒子喷注所占的比例。灵敏度用来衡量网络正确判别胶子(夸克)的能力，是将测试样本中真实的胶子(夸克)喷注正确地判定为胶子(夸克)喷注的比例，其灵敏度越高，漏判率就越低。特异性用来衡量网络正确判别不是胶子(夸克)的能力，是将测试样本中不是胶子(夸克)喷注正确地判定为不是胶子(夸克)喷注的比例，其特异性越高，误判率就越低。

3.2 实验数据的判别分析

为了评价新算法的性能，我们通过计算机模拟实验与原基于负梯度法的 BP 算法进行了比较。实验中选用 2500 个胶子喷注和 2500 个夸克喷注组成训练样本进行训练，另取另外相同比例的 1000 个胶子喷注和 1000 个夸克喷注组成测试样本进行测试。由于神经网络的输出值介于 $(0, 1)$ 范围内，所以取输出中值为收敛度，即 $\sigma = 0.5$ ，实验也发现这样能够更加精确地鉴别未知胶子和夸克喷注样本。

针对 3-2-2 和 3-6-2 的 BP 网络结构，实验中分别采用负梯度算法和共轭梯度优化算法对训练样本各训练 1000 次，在相同的条件下，不同的算法得到的结果如表 1 所列。从表 1 中我们可以发现共轭梯度优化算法在相同条件下，对于不同的网络结构不仅判定效率比负梯度法分别高出 1.6% 和 2.4%，而且时间也分别节省 5 s 和 4 s，可见，采用优化算法的神经网络鉴别粒子喷注的能力较强，同时网络计算耗时也较少。其中，当网络结构为 3-2-2 时，优化算法计算得到的判定效率总是 1，这就说明该优化算法在实验中不存在无法判别的粒子喷注，被判别为胶子喷注中错误的粒子喷注就是夸克喷注，而被判别为夸克的喷注中错误的粒子喷注就是胶子喷注。另外，当网络结构为 3-2-2 时，比较不同的算法，经负梯度算法计算得到胶子灵敏度虽然高出了 1% 左右，但由于胶子特异性降低了 10% 以上，从而得到实验结果误判率相对

较高。同样负梯度算法计算得到夸克的特异性虽然高出了1.3%左右,但由于夸克的灵敏度降低了9.8%以上,从而得到实验结果漏判率相对较高。而经优

化算法计算得到的粒子喷注的灵敏度和特异性都高于89%,这就意味着得到的实验结果漏判率和误判率均低,算法稳定性较高。

表1 运用共轭梯度优化算法和负梯度算法对高能物理夸克和胶子鉴别的结果

方法	网络结构	时间/s	判定效率/%	夸克灵敏度/%	胶子灵敏度/%	夸克特异性/%	胶子特异性/%
负梯度算法	3-6-2	91	0.973	0.776	0.995	0.996	0.881
共轭梯度优化算法	3-6-2	86	0.989	0.874	0.987	0.988	0.891
负梯度算法	3-2-2	87	0.976	0.816	0.988	0.993	0.733
共轭梯度优化算法	3-2-2	83	1	0.947	0.978	0.978	0.947

比较不同的网络结构,当网络结构为3-2-2时,对于不同的算法均能得到理想的实验结果。可见,对于具体问题不是网络越复杂越好,相反,对于那些训练样本较少的情况,较少的隐藏层节点数就可以得到很好的结果,而且网络计算花费的时间也较少。

共轭梯度优化算法采用共轭方向进行搜索,并通过步长优化得到最佳步长,使目标函数快速收敛到极小值。实验表明,我们设计的优化算法不仅增强了神经网络的学习效率,而且也增强了网络的有效性。

另外,将该优化算法应用于其它判别领域中,同样也可以得到令人满意的分类结果。例如,采用优化算法对Iris数据集进行分类。Iris数据集是UCI数据

库^[18]中使用频率最高的数据集,在分类和模式识别研究中被经常采用。此数据集以鸢尾花的特征作为数据来源,包含150个数据样本,分为3类,每类50个数据样本,且服从正态分布,每个数据样本包含4个特征属性。在这里,为了和上述3-2-2网络结构相对应,我们选取了前两个类的所有数据(共100个数据样本)进行实验,其中取每个类的前25个数据组成训练样本,后25个数据组成测试样本。在相同收敛度($\sigma = 0.5$)下,不同的算法得到的结果如表2所列。可以看出,对于Iris数据集的分类结果,我们设计的共轭梯度优化算法比负梯度法在灵敏度和特异性上均有较大提高。

表2 运用共轭梯度优化算法和负梯度算法对Iris数据判别的结果

方法	网络结构	时间/s	判定效率/%	夸克灵敏度/%	胶子灵敏度/%	夸克特异性/%	胶子特异性/%
负梯度算法	3-6-2	2	1	0.92	1	0.92	0.925
共轭梯度优化算法	3-6-2	1.6	1	0.96	1	0.96	0.961

4 结束语

本文利用共轭梯度法和二次插值法设计出了一种新的BP算法,该算法有效地避免了传统BP算法靠人为经验确定步长的缺陷,大大提高了神经网络快速稳定收敛的能力,步长优化耗时也较少。在物理夸克胶子喷注的鉴别和Iris数据判别中,可以发现采用共轭梯度优化算法,通过设定相应的网络模型,能使网络快速地收敛到目标函数的极小值,获得有效的鉴别效果,不仅增强了网络的有效性,同时也增强了网络的通用性。

参考文献:

- [1] ZHU Yongsheng. Experimental Data Multivariate Statistical Analysis[M]. Beijing: Science Press, 2009. (in chinese) (朱永生. 实验数据多元统计分析[M]. 北京: 科学出版社, 2009.)
- [2] RISH I. An Empirical Study of the Naive Bayes Classifier[C]//In proceedings of IJCAI-01 workshop on empirical methods in artificial intelligence, 2011.
- [3] QUINLAN J R. Induction of Decision Trees. Machine Learning, 1986(1): 81.
- [4] YANG Xiaomin, HE Xiaohai. Journal of Sichuan University, 2006, 43(5): 1043. (in chinese)

- (杨晓敏, 何小海. 四川大学学报, 2006, **43**(5): 1043.)
- [5] BELHUMEUR P N, NESPANHA J P, DVID J K. IEEE Trans on Pamt, 1997, **19**(7): 711.
- [6] DENBY B. Computer Physics Communications, 1988, **49**(3): 429.
- [7] ZHOU Yu. Journal of HuaZhong University of Science and Technology, 2012, **17**: 19. (in chinese)
(周玉. 华中科技大学学报, 2012, **17**: 19.)
- [8] SUN Gongxin, DAI Changjiang, DAI Guiling. Nuclear Electronics and Detection Technology, 2003, **1**: 23. (in chinese)
(孙功星, 戴长江, 戴贵亮. 电子学与探测技术, **1**: 23.)
- [9] ZHANG Kunshi, LIU Lianshou. China Physics C, 2004, **1141**: 1145.
- [10] LONNBAD L, PETERSON C, ROGNVALDSSOM T. Physical Review Letters, 1990, **65**: 1321.
- [11] CSABAI I, CZAKÓ F, FODOR Z. Phys Rev D, 1991, **44**: R1905.
- [12] RIPLEY B D. Pattern Recognition and Neural Networks[M]. Beijing: Posts & Telecom Press, 2009: 73. (in Chinese)
- (里普利. 模式识别与神经网络[M]. 北京: 人民邮电出版社, 2009: 73.)
- [13] HAYKIN Simon. Neural Networks: A Comprehensive Foundation[M]. SHIWEI Y E. 2nd ed. Beijing, China Machine Press, 2004: 17. (in Chinese)
(海金. 神经网络原理[M]. 叶世伟, 译. 第二版. 北京: 机械工业出版社, 2004: 17.)
- [14] HAGAN M T. Neural Network Design[M]. DAI kui. Beijing: China Machine Press, 2002: 40. (in Chinese)
(哈根. 神经网络设计[M]. 戴葵, 译. 北京: 机械工业出版社, 2002: 40.)
- [15] FAN Lin, DUAN Fujian. Journal of Guilin University of El Ectronic Technology, 2007, **6**: 27. (in chinese)
(范林, 段复建. 桂林电子科技大学学报, 2007, **6**: 27.)
- [16] JACOBS R A. Neural Networks, 1988, **1**(4): 295.
- [17] HANSON G. Phys Rev Lett, 1975, **35**: 1609.
- [18] FRANK A, ASUNCION A. UCI Machine Learning Repository[EB/OL]. [2013-07-10]. [http://archive.ics.uci.edu/ml].

Application of the BP Neural Network Based on Conjugate Gradient Optimization Algorithm in the Identification of High Energy Particles and Other Fields

WANG Shuwang¹, LU Yonggang¹, CHEN Xurong^{2, 3}

(1. School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China;

2. Institute of Modern Physics, Chinese Academy of Sciences, Lanzhou 730000, China;

3. Lanzhou hadron physics and CSR Physics Research Center, Lanzhou 730000, China)

Abstract: Artificial neural network methods have been introduced in high energy physics experiments and have been widely applied to the identification of the quark-gluon injection, electronic hadron discrimination, top quark, and the Higgs particle searching and so on. This paper introduces a modified conjugate gradient optimization algorithm, which is applied to the identification of high-energy particles. In the application, the algorithm can obtain optimal step size in the search direction for minimizing the objective function, and can overcome the local vibration problem, so that the fast convergence of the objective function is obtained and the stability of the algorithm is improved. The analysis of experimental data shows that our new BP neural network algorithm can effectively improve the identification of particles in high energy physics.

Key words: BP neural network; conjugate gradient; step optimization; particle identification

Received date: 11 Oct. 2013; **Revised date:** 8 Jan. 2014

Foundation item: National Natural Science Foundation of China (11005045, 11175220, 61272213); Hundred Talents Program of Chinese Academy of Sciences(Y101020BR0)

Corresponding author: CHEN Xurong, E-mail: xchen@impcas.ac.cn.

<http://www.npr.ac.cn>