

神经网络在 $\tau \rightarrow \rho\nu_\tau$ 极化分析中的应用

张子平

(中国科学技术大学近代物理系 合肥 230026)

王贻芳

(Lab. for Nuclear Science, MIT, USA)

V. Innocente

(CERN and INFN-Sezione di Napoli, Italy)

1993年5月3日收到

摘 要

在 LEP/L3 的 $\tau \rightarrow \rho\nu_\tau$ 极化物理分析中,用人工神经网络法选择事例,获得了比较高的选择效率。详细研究表明,网络本身不会引入明显的系统偏差。使用在 Z^0 能区采集到的 8977 个 $\tau^+\tau^-$ 对事例,得到 $\tau \rightarrow \rho\nu_\tau$ 道的极化率 $P_\tau = -0.129 \pm 0.050 \pm 0.050$ 。本文介绍在这一工作中使用的神经网络训练方法和技巧。

关键词 极化,神经网络,训练方法。

1 引 言

非极化束 $e^+e^- \rightarrow Z^0 \rightarrow f^+f^-$ 道的极化不对称性

$$P_f = \frac{\sigma(f_L) - \sigma(f_R)}{\sigma(f_L) + \sigma(f_R)} \quad (1)$$

可用来测量弱中性流的宇称破坏,精确检验标准模型。这里 L 和 R 指 f 的左和右螺旋性。实验上可行的道是 τ 轻子道^[1,2]。由于 P_τ 对质心系总能量 \sqrt{s} 的变化不敏感,不仅可给出 $\sin^2\theta_W$ 的精确测量,而且可测得 ν_τ 和 a_τ 的相对符号。对轻子衰变道 $\tau \rightarrow e\nu$, $\mu\nu$, 以及半轻子衰变道 $\tau \rightarrow \pi\nu$, P_τ 可通过拟合末态带电粒子的能谱而测得。而对于具有很大分支比,在 τ 的极化测量中很重要的半轻子衰变道 $\tau \rightarrow \rho\nu$, 记 $x_\rho = E_\rho/E_\tau$, 树图下的能谱可表为

$$\frac{1}{N} \cdot \frac{dN}{dx_\rho} = 1 + \alpha_\rho P_\tau (2x_\rho - 1), \quad (2)$$

这里,依赖于 ρ 质量的灵敏常数 $\alpha_\rho \approx 0.46$, 因而压制了 ρ 能谱对 P_τ 测量的灵敏度,必需进一步研究 ρ 衰变产生的 π 介子末态的分布来获得对 P_τ 的灵敏测量。若定义 θ^* 角为在 τ 的静止坐标系中 ρ 和 τ 的飞行方向之间的夹角, ψ^* 角为在 ρ 的静止坐标系中带电 π 介子和 ρ 的飞行方向之间的夹角,拟合 $\cos\theta^*$, $\cos\psi^*$ 的两维分布可以获得对 P_τ 的灵

敏测量。然而困难在于,由于飞行 τ 粒子的洛仑兹效应,使得末态带电 π 介子和中性 π 介子之间的夹角很小,在L3探测器中难以将它们区分开来,因此用通常的方法把 $\tau \rightarrow \rho\nu \rightarrow \pi\pi^0\nu$ 和别的一些反应道,诸如 $\tau \rightarrow \pi\nu, \tau \rightarrow a_1\nu \rightarrow \pi\pi^0\pi^0\nu$ 和 $\tau \rightarrow \pi\pi^0\pi^0\pi^0\nu$ 等鉴别开来相当困难。

近年来神经网络法被广泛应用于高能物理研究领域,诸如粒子的鉴别,标记以及径迹重建,在线触发系统等许多方面^[3-6]。这里介绍我们在 $\tau \rightarrow \rho\nu \rightarrow \pi\pi^0\nu$ 事例选取中使用的一个前馈式(feed-forward)神经网络系统,具体地讨论在这一工作中使用的训练方法和技巧。最后给出用于该反应道的极化分析所得到的结果。

2 L3探测器和事例预选

L3探测器主要包含中心径迹探测器,高分辨BGO电磁量能器,强子量能器,闪烁计数器,精密的大型 μ 子漂移室系统以及亮度监测器等几个部分,组装在一个直径25m、场强0.5T的磁场中。中心径迹探测器由两层多丝正比室(z 室)包围的圆柱型时间扩展室(TEC)组成,TEC又分内、外室,分别含12和24个单元,动量分辨为 $\sigma(1/p_T) = 0.022\text{GeV}$,位置分辨在BGO表面处垂直于束流方向的平面内为0.5mm。电磁量能器由桶部7680根、端部 2×1536 根BGO晶体组成。强子量能器由铀板和多丝正比室夹层组成。对该分析,量能器的分辨率对能量大于1.5GeV的电子和 τ 好于2%,对带电 π 介子为 $55\%/\sqrt{E(\text{GeV})} + 8\%$ 。 μ 子探测器对 $p_T = 45\text{GeV}$ 的带电粒子动量分辨为2.8%。关于L3探测器及其性能的详细描述可见文献[7]。

根据TEC测得的带电粒子径迹,电磁量能器和强子量能器中能量沉积的簇团,以及 μ 子漂移室测得的 μ 子径迹等信息,就可以重建事例,给出事例中最小可分辨的物理实体,称为SRC's,对应于一个单独的粒子或一组不可分的粒子。这些SRC's又可按照一定的算法^[8]构成喷注。基于探测器提供的所有物理信息,一套物理变量的截断值被用来进行 $Z^0 \rightarrow \tau^+\tau^-$ 事例的预选,在探测器的桶部区域($|\cos\theta| < 0.7$),效率达84.9%,本底水平1.8%,详见文献[9,10]。在我们的分析中另外要求只选取单叉的喷注事例,即每一喷注中有且只有一个好的带电粒子径迹,对应于 τ 衰变末态中的带电粒子,其动量大于 $0.3\text{GeV}/c$ 。

3 网络结构

我们设计的神经网络结构如图1所示,由一个输入层,一个隐藏层和一个输出层组成。输入层和隐藏层分别设计为18个网点,输出层为一个网点。输入层的网点对应于 τ 事例的18个物理变量,它们的名称和意义如下:

- 1) E_{BGO} , BGO电磁量能器中沉积的总能量。
- 2) E_{had} 强子量能器中沉积的总能量。
- 3) p_{TEC} 中心时间扩展径迹室量得的带电径迹动量。
- 4, 5, 6) $\Sigma_{\text{BGO}}^1, \Sigma_{\text{BGO}}^2, \Sigma_{\text{BGO}}^3$ 分别为BGO电磁量能器中沉积能量最大的簇团中两

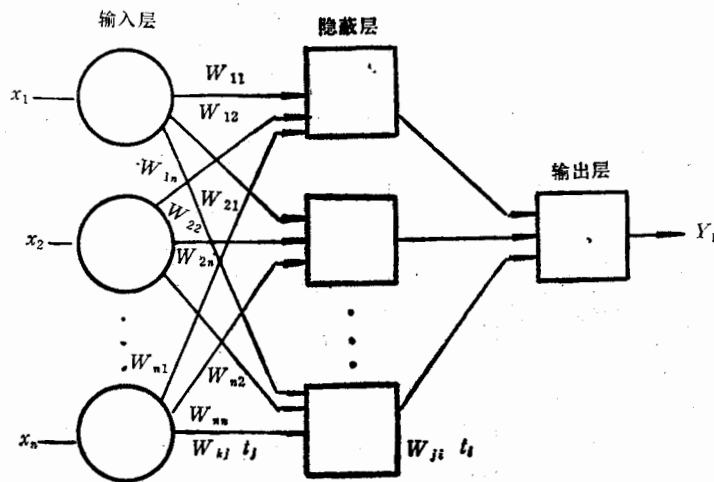


图 1 前馈式神经网络结构

个能量最高的 BGO 晶体中能量之和,以及围绕能量最高的晶体的 9 个及 25 个晶体中的能量之和。

7,8) E_{src}^1, E_{src}^2 量能器中能量最大和次大的 SRC 中的能量。

9,10) E_b^1, E_b^2 , BGO 电磁量能器中沉积能量最大和次大的簇团中的能量。

11,12) E_{nb}^1, E_{nb}^2 , BGO 电磁量能器中沉积能量最大和次大的中性簇团中的能量,所谓中性是指在 $R-\varphi$ 平面 15mrad 内没有带电径迹指向该能量簇团。

13,14) W_θ, W_φ 最大的能量簇团在沿束流方向和垂直于束流方向的宽度。

15) N_{nb} 喷注中中性能量簇团的数目。

16) N_{src} 喷注中 SRC 的数目。

17) R_s , BGO 中能量簇射形状的长轴和短轴之比。

18) Δ_{tb} , TEC 中的径迹和 BGO 中最大能量簇团的重心之间在 $R-\varphi$ 平面内 φ 角之差。

图 2 是经探测器模拟和重建后的 τ 模拟数据给出的这些物理量中的一部分对不同衰变道的分布。

网络的隐藏层和输出层分别通过权重 (W) 和阈值 (t) 与前一层的输出相联系,例如,由输入层到隐藏层,依图 1 的符号记法,

$$a_j = \sum W_{kj} x_k + t_j, \quad (3)$$

将一个 S 形函数

$$g(x) = 1/(1 + \exp\{-x/T\}) \quad (4)$$

作用到 a_j 上,得到隐藏层第 i 个网点的输出为

$$h_j = g(a_j). \quad (5)$$

取温度常数 $T = 1$,不同的 T 值将改变 S 形函数的形状。S 形函数将输出 h_j 压缩在 $[0,1]$ 区间,作为下一层的输入。由隐藏层到输出层的操作和前面完全相同。

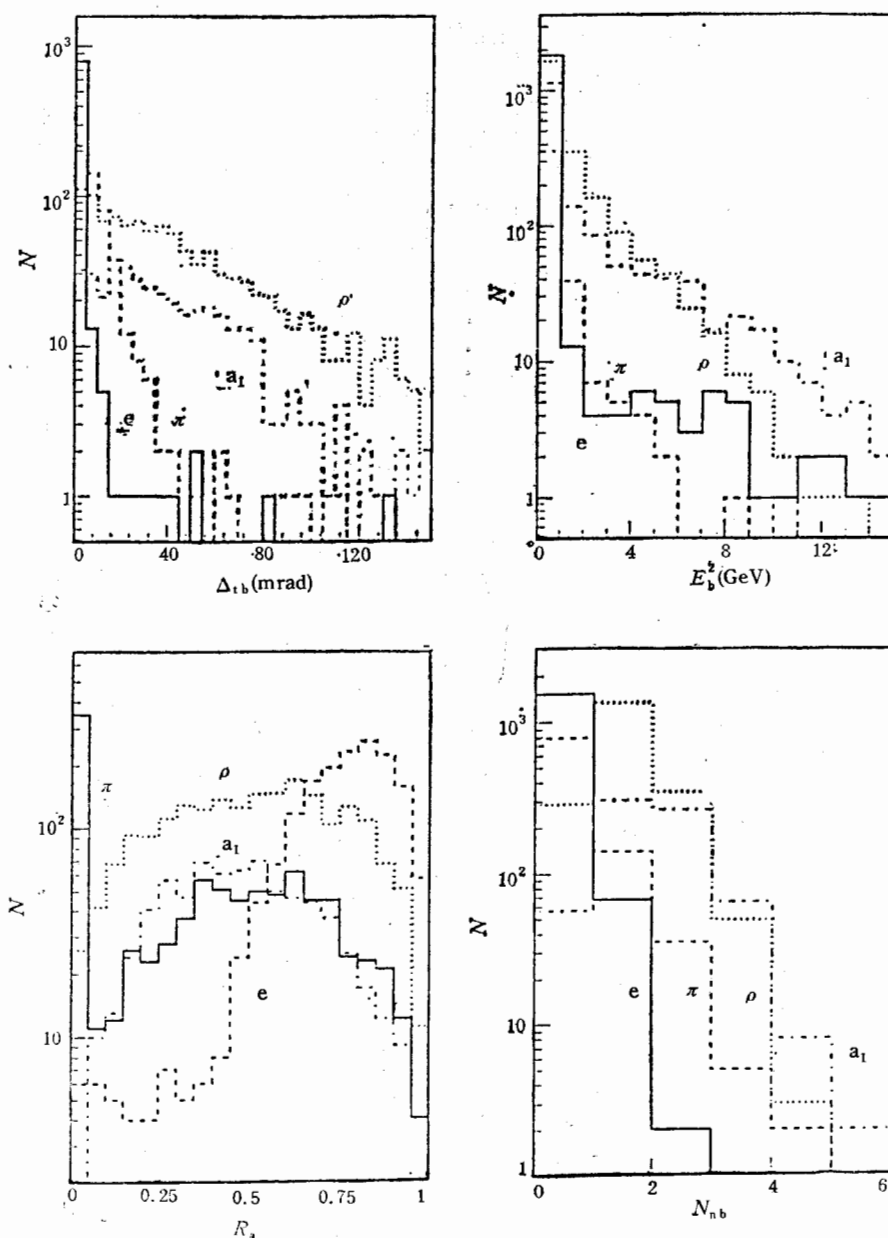


图2 网络的一部分 τ 事例输入变量对不同衰变道的分布(纵坐标 N 为事例数)

4 网络的训练和检验

记由输出层网点的最后输出为 Y , 则训练网络实质上就是选取 (W) 和 (t) 的最佳值, 使得下面的误差函数取极小值.

$$E = 0.5 \cdot \sum_P \sum_i (Y_i(P) - A_i(P))^2, \quad (6)$$

这里的 P 指事例样本, i 是输出层网点数, 对我们的情况 $i = 1$, A 是对输出的设计值, 约定赋值

$$A = \begin{cases} 1, & \text{对信号事例} \\ 0, & \text{对本底事例.} \end{cases} \quad (7)$$

使用 KORALZ3.18^[11] 产生程序产生的 $e^+e^- \rightarrow \tau^+\tau^-$ 事例, 经过 L3 探测器模拟和重建, 选取在探测器桶部 ($|\cos\theta| < 0.7$) 区域的经过预选的 4 万个模拟事例作为网络的输入, 采用误差向后传播法 (BP)^[12] 对网络进行训练. 首先对所有的权重 (W) 和阈值 (t) 在 $[-0.1, 0.1]$ 区间随机地赋以初值, 然后用下面的算法来调整它们达到最佳值, 使得误差函数 E 取极小.

1) 由输出层到隐藏层. 为使误差函数 E 逐渐减小, W_{ji} 应沿 E 的负梯度方向变化, 即应有

$$\Delta W_{ji} = -\eta \cdot \frac{\partial E}{\partial W_{ji}}, \quad (8)$$

$$\frac{\partial E}{\partial W_{ji}} = \delta_i g'(a_i) h_j, \quad (9)$$

这里定义了 $\delta_i = (Y_i - A_i)$, (8) 式可改写为

$$\Delta W_{ji} = -\eta \delta_i g'(a_i) h_j + \alpha \Delta W_{ji}^{\text{old}}, \quad (10)$$

α 项称为动量项, 是为了抑制振荡而加入的, $\Delta W_{ji}^{\text{old}}$ 指上一次循环得到的值, η 称为学习强度. 同样对 t_j 有

$$\Delta t_j = -\eta \delta_j g'(a_j) + \alpha \Delta t_j^{\text{old}}. \quad (11)$$

2) 由隐藏层到输入层与前述相似, 只是由于这时

$$\frac{\partial E}{\partial W_{ki}} = \sum \delta_i g'(a_i) W_{ji} g'(a_j) x_k = \delta_i' g'(a_j) x_k, \quad (12)$$

因而只需把(10)和(11)式中的 δ_i 换成 $\delta_i' = \sum \delta_i g'(a_i) W_{ji}$ 即可.

在训练中取 $\alpha = 0.7$, η 开始时取 0.5, 随着训练循环次数的增加逐渐减小到 0.001. 每次循环中随机地选取一个信号事例和一个本底事例作为网络的输入, 调整所有的权重和阈值, 它们的变化量 ΔW 和 Δt 则在每 5 次循环后更新一次. 训练过程一般取 250 到 300 万个循环, 耗 Apollo DN10000 CPU 大约一个小时.

图 3 给出了取 $\tau \rightarrow \rho\nu$ 作为信号, 所有别的 τ 衰变事例作为本底得到的训练输出 Y 值的分布, 信号事例和本底事例被明显地分开, 兼顾事例的选择效率和纯度, 可以给定 Y 的截断值 Y_{cut} , 称 $Y > Y_{\text{cut}}$ 的事例为候选信号事例. 这里选择效率指的是在 $|\cos\theta| < 0.7$ 区域选择到的真信号事例占应有真信号事例的百分比, 纯度则是指选择到的真信号事例占候选信号事例的百分比, 也即 1 减去本底水平. 图 4 给出了它们作为 Y_{cut} 函数的分布.

网络一旦训练成功后, 所有的权重和阈值都固定下来, 不再改变. 这时网络的作用就象一个函数型的“黑匣子”, 对于一个给定的输入, 就产生一个确定的输出. 由于用于训练的模拟数据样本的统计有限性, 训练过程中样本内部复杂的自调整和补偿效应, 训练样本给出的结果必然较实际情况要好一些. 为反映网络的真实性能, 使用另外一套 14 万个完

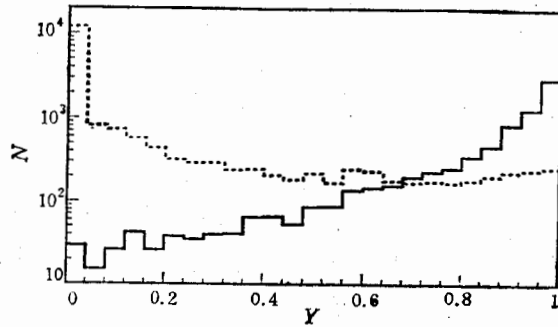


图3 网络训练结果给出的 $\tau \rightarrow \rho\nu$ 信号和本底的 Y 值分布
本底, —信号.

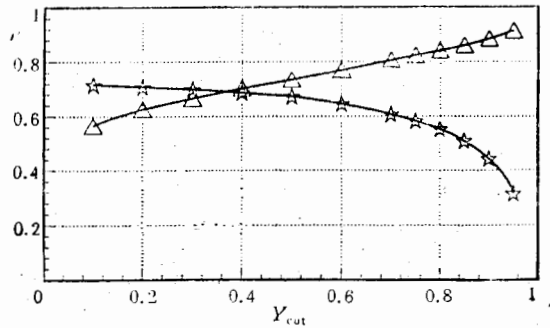


图4 网络训练结果给出的 $\tau \rightarrow \rho\nu$ 事例选择效率(星号)和纯度(三角形)随 Y_{cut} 的变化

全统计独立的相同的 τ M.C. 事例样本作为网络函数“黑匣子”的输入, 对其进行检验. 表1给出了 $Y_{\text{cut}} = 0.8$ 时的检验结果和训练结果的比较.

5 结果和讨论

表1给出, 我们设计的前馈式神经网络, 在 L3 实验中用以选择 $\tau \rightarrow \rho\nu$ 事例可以达到 54% 的效率和 83% 的纯度. 图5给出了 τ 事例 M.C. 和实验数据的输出 Y 值分布, 两者符合得很好. 取 $Y_{\text{cut}} = 0.8$, 对选择到的候选事例构造出 $\pi\pi^0$ 的不变质量, 峰在 $800\text{MeV}/c$ 处, 与 ρ 的质量 $768.3 \pm 0.5\text{MeV}/c$ 相符. 利用 L3 采集到的 8977 个 $\tau^+\tau^-$ 对事例, 取 $Y_{\text{cut}} = 0.8$, 用 M.C. 事例减除本底, 对最后选择到的事例, 用最大似然法拟合 $\cos\theta^*$ 和 $\cos\psi^*$ 的两维分布, 每维 20 等份, 给出分析结果 $P_r = -0.129 \pm 0.050 \pm 0.050^{[10]}$.

表1 训练结果和检验结果的比较

	选择效率(%)	本底水平(%)	M.C. 样本数
训 练	55.2	16.1	40000
检 验	53.8	17.2	140000

所有 τ 事例的 18 个输入物理量, 在输入网络前, M.C. 和实验数据之间都符合得很好 (图 6). 经过网络的操作后, 选取不同的 Y_{cut} 值得到的 $\tau \rightarrow \rho\nu$ 候选事例, M.C. 和实验数据也都很好地相符 (图 7). 作为相互校验, 把相同的网络用以筛选 $\tau \rightarrow e\nu\nu$ 和 $\tau \rightarrow \pi\nu$ 道事例, 也都得到了理想的选择效率和本底水平. 仔细研究了网络本身在各种想得到的情况下可能带来的误差, 诸如输入变量大小和个数的变化, 能动量的分辨率和刻度误差, 不同的 Y_{cut} 值, 效率修正, 各变量间的相关性等各种可能因素的影响, 没有发现网络本身有引入明显的系统偏差的可能性^[12]. 这种网络具有很强的容错功能. 我们虽然不能证明网络在训练中没有陷入局域极小值, 但是做了多种检验, 没有得到更好的结果, 相信训练

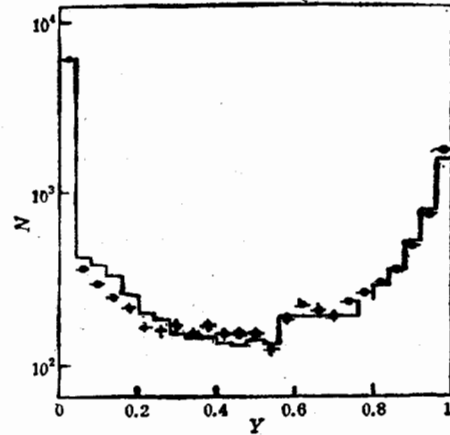


图 5 网络检验输出的 τ 事例实验和 M.C. 数据比较
实线是 M.C., 圆点是实验点.

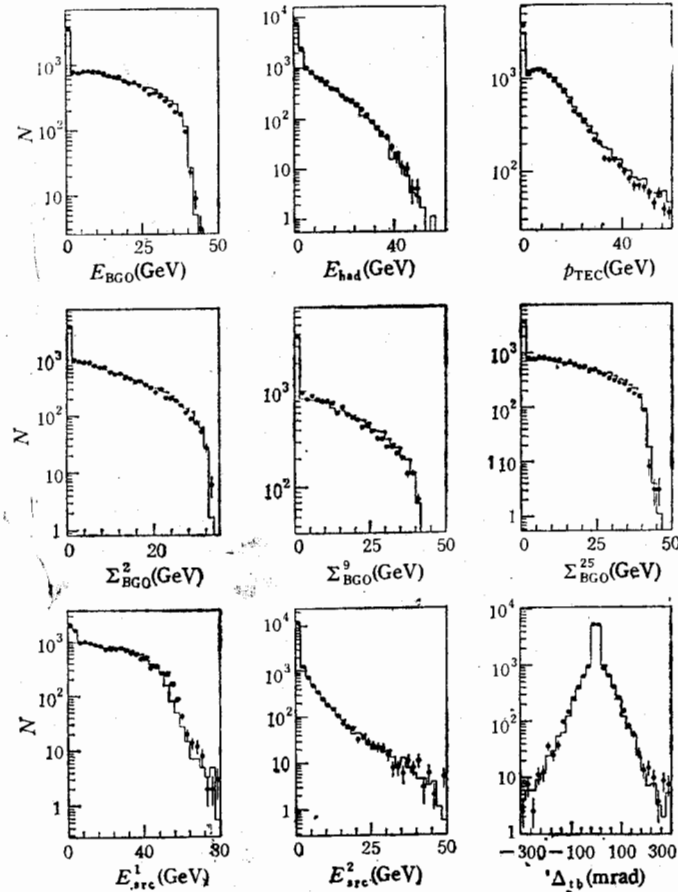


图 6 网络的一部分 τ 事例输入变量的实验和 M.C. 数据比较
实线是 M.C., 圆点是实验点.

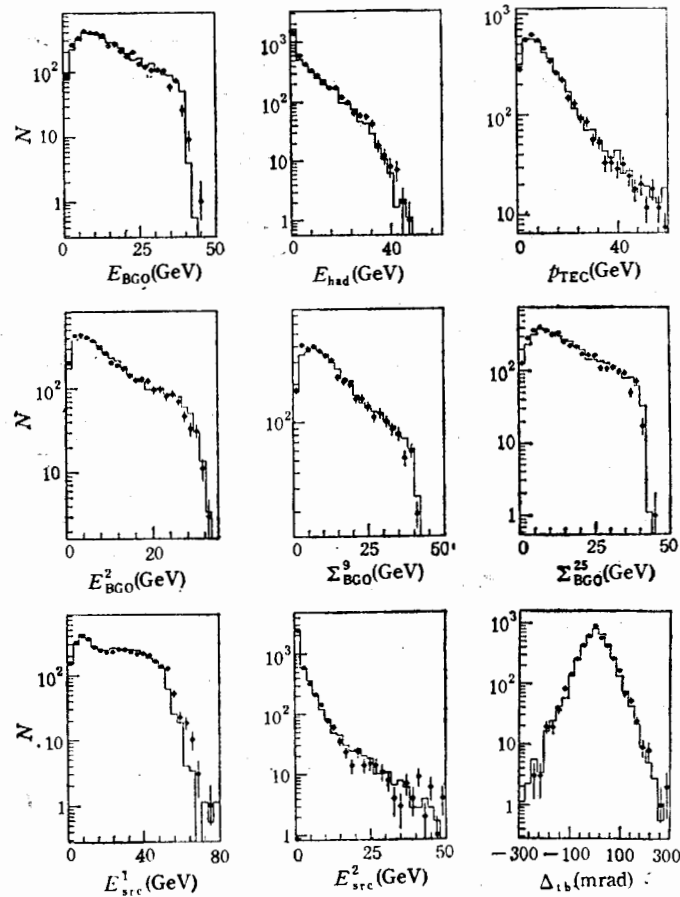


图7 经网络选择后的一部分 $\tau \rightarrow \rho\nu$ 候选事例变量的实验和 M.C. 数据比较
实线是 M.C., 圆点是实验点。

陷入局域极小值的可能性很小。

作者感谢在这一工作中 L3 合作组同行们给以的支持和帮助, 特别是陈和生教授和 R. Mount 教授等的一些有益的建议和讨论。本文第一作者也感谢西欧核子研究中心为其在那里工作期间提供的支持和便利。

参 考 文 献

- [1] A. Rougé, Tau decays as polarization analysis, Talk at the Workshop on Tau Physics, Orsay, France, Sep. 24—27, 1990.
- [2] K. Hagiwara et al., *Phys. Lett.*, **B235**(1990) 198.
- [3] For an introduction, see, for example, P. Wasserman, *Neural Computing: Theory and Practice* (Van Nostrand Reinhold, N.Y., 1989).
- [4] L. Lönnblad et al., LU TP90-8 and *Nucl. Phys.*, **B349**(1991)675.
- [5] L. Bellantoni et al., Using Neural Network with Jet Shape to Identify b Jets in e^+e^- Interactions, CERN-PPE/91-80.
- [6] Proceedings of the Inter. Conf. on Computing in High Energy Physics'92, Annecy, France,

- CERN 92-07, P639-677, and References There.
- [7] L3 Collaboration, B. Adeva et al., *Nucl. Instr. and Meth.*, **A289**(1990)35.
- [8] O. Adriani et al., *Nucl. Instr. and Meth.*, **A302**(1991)53.
- [9] L3 Collaboration, B. Adeva et al., *Z. Phys.*, **C51**(1991)179.
- [10] L3 Collaboration, B. Adeva et al., *Phys. Lett.*, **B294**(1992)466.
- [11] S. Jadach et al., CERN Report CERN-89-08, Vol.III, (1989)69.
- [12] V. Innocente, Y.F. Wang and Z.P. Zhang, *Nucl. Instr. and Meth.*, **A323**(1992)647.

Application of Neural Network in $\tau \rightarrow \rho\nu_\tau$ Polarization Analysis

Zhang Ziping

(*Uni. of Science and Technology of China, Hefei 230026*)

Wang Yifang

(*Lab. for Nuclear Science, MIT, USA*)

V. Innocente

(*CERN and INFN-Sezione di Napoli, Italy*)

Received on May 3, 1993

Abstract

An artificial neural network was built to select events in the $\tau \rightarrow \rho\nu_\tau$ polarization analysis at LEP/L3, much better selection efficiency has been achieved. Detailed studies show that no systematic errors or bias have been introduced by the application of neural network. A polarization of $P_\tau = -0.129 \pm 0.050 \pm 0.050$ for this channel was obtained by using a sample of 8977 $\tau^+\tau^-$ pairs collected near the peak of Z^0 resonance. The neural network training method and some details are described.

Key words polarization, neural network, training method.