

在 BES 实验中用神经网络法 鉴别 e, μ, π^*

张子平 陈宏芳 叶树伟 赵家伟

(中国科技大学近代物理系 合肥 230026)

1996-03-18 收稿

摘 要

研究了两种不同的神经网络法, BP 网络和 LVQ 网络, 对北京谱仪 (BES) 实验中 e, μ, π 粒子的鉴别, 取得了较常规方法要好的结果. 用于训练和检验的 μ 子样本来自宇宙线事例, e 和 π 粒子则是由真实实验数据精选的, 虽然样本本身具有非均匀的动量谱, 但 BP 网络的检验结果给出的粒子选择效率在整个动量区间却仍然具有相当均匀的分布, LVQ 网络稍逊之. 至少是在我们研究的这一课题中证明了 BP 网络的模式识别功能要优于 LVQ 网络.

关键词 BES 实验, 粒子鉴别, 神经网络法.

1 引 言

e, μ, π 粒子的鉴别在 BES 实验的离线分析中具有十分重要的意义. 用通常判别法, 即设置许多物理量的截断条件, 对这些粒子的选择效率和本底压制水平都不能令人满意, 促使人们设法寻找别的更有效的事例判选方法^[1].

近些年来, 作为一种高效和强有力的模式识别方法, 神经网络法已经被越来越多地应用于高能物理实验的数据分析中, 取得了相当令人满意的物理结果^[2-6]. 为对 BES 实验中的 e, μ, π 粒子进行高效鉴别, 我们尝试使用了两种完全不同的神经网络法, BP(back-propagation) 网络和 Kohonen 的 LVQ (Learning Vector Quantization) 网络, 在如下诸方面进行了较细致的探讨.

(1) 通常的作法是用蒙特卡罗模拟事例对网络进行训练和检验, 以确定网络的判选能力. 而在 BES 实验中, 由于模拟数据不准确, 无法使用, 因而不得不使用经严格精选的真实实验事例和宇宙线事例. 这样做面临的一个问题是, 事例的动量谱不均匀, 那么能否保证训练结果给出的选择效率随动量的变化均匀呢?

(2) 在完全相同的应用条件下, 对 BP 网络和 LVQ 网络进行比较, 以确定在实际应

* 中国科学院高能物理研究所开放实验室和国家自然科学基金资助.

用中哪种方法更有效.

(3) 定量地确定 e, μ, π 事例的选择效率和本底污染水平.

2 事例样本的筛选

用于网络训练和检验的 μ 子样本是由 BES 合作组的同事们提供的非常干净的宇宙线数据, e, π 样本则是我们由真实实验数据中严格精选的, 筛选条件如下:

电子样本由 $ee\gamma$ 事例中选取. 基本思想是认准动量较高的那支径迹为电子, 则把低动量未加截断条件的那支径迹作为电子样本. 对高动量一支的判选条件为: 在 $|\cos\theta| \leq 0.80$ 的 BES 覆盖区间, 要求动量 $p \geq 1.0 \text{ GeV}/c$; 由飞行时间 (TOF) 和动量 p 计算得到的速度 $0.8 \leq v/c \leq 1.4$; μ 子探测器没有计数; dE/dx 的测量值与电子径迹期待值的偏离小于两倍标准偏差; 桶部簇射计数器 (BSC) 中的沉积能量 E_{BSC} 与动量 p 之比满足 $0.6 \leq E_{\text{BSC}}/p \leq 2.0$; BSC 前 5 层中的粒子击中数之和与 12 之差的绝对值大于或等于 3.

π 样本是从 $J/\psi \rightarrow \rho\pi, \omega\pi\pi, \omega 4\pi$ 等过程中选取的. 要求事例满足运动学能、动量守恒条件; 中性径迹数为 2—6; 总电荷等于 0; TOF 的最大值与最小值之差的绝对值小于 4.0 ns; 径迹运动学拟合的 $\chi^2 < 25$, 组成 π^0 的两个 γ 光子的不变质量截断条件为 $|m_{2\gamma} - 0.135| < 0.05 \text{ GeV}/c^2$; 此外三个衰变道各自的判选条件为:

对 $J/\psi \rightarrow \rho\pi$ 道, 要求带电径迹数等于 2, 组成 ρ 的两个 π 的不变质量 $|m_{2\pi} - 0.770| < 0.2 \text{ GeV}/c^2$.

对 $J/\psi \rightarrow \omega 2\pi, \omega \rightarrow \pi^+ \pi^- \pi^0$ 道, 要求带电径迹数等于 4, ω 的不变质量截断为 $|m_{3\pi} - 0.782| < 0.05 \text{ GeV}/c^2$.

对 $J/\psi \rightarrow \omega 4\pi$ 道, 要求带电径迹数等于 6, ω 的不变质量截断同上.

3 BP 网络的训练和检验

3.1 BP 网络的结构和训练

研究比较了两种训练与检验方案, 一种是使用整个动量区间的样本进行总体操作, 另一种是考虑到只有高动量的 μ 子及部分 π 径迹才有可能打到 μ 子计数器, 为能充分利用 μ 子计数器中的信息对高动量粒子进行最有效的鉴别, 把 $p \geq 0.5 \text{ GeV}/c$ 和 $p < 0.5 \text{ GeV}/c$ 的两类粒子径迹分别进行训练和检验. 结果显示两种方案得到的结果没有很明显的差别.

在认真研究了三类粒子在探测器内的物理响应的基础上, 对 $p < 0.5 \text{ GeV}/c$ 的粒子选择了如下的 23 个物理变量:

(1) p , 带电径迹的动量;

(2—4) x_{se}, x_{su}, x_{sp} , 假定径迹为 e, μ 或 π 时, 电离能量损失 dE/dx 与其期待值的偏离;

(5) TOF, 飞行时间;

(6) E_{BSC} , BSC 中的能量沉积;

(7—12) $Hit(i), i=1, \dots, 6$, BSC 第一层到第六层各层中的击中数;

(13—18) $Se(i), i=1, \dots, 6$, dE/dx 在 BSC 各层中的脉冲高度;

(19, 20) $|\theta_{MDC} - \theta_{BSC}|, |\phi_{MDC} - \phi_{BSC}|$, 由主漂移室(MDC)决定的方向和 BSC 第一层击中点决定的方向之间极角和方位角之差的绝对值.

(21) N_h^c , BSC 总的击中单元数;

(22, 23) N_h^{max}, L_{BSC} , 在 BSC 的第 L_{BSC} 层有最多击中单元数 N_h^{max} .

对 $p \geq 0.5 \text{ GeV}/c$ 的粒子增加了下面与 μ 子计数器信息有关的 8 个物理量:

(24—29) $D_R(i), D_z(i), i=1, 2, 3$, μ 子计数器各层的击中点和由主漂移室到该层的外推点之间在垂直束流的 R 方向和沿束流的 z 方向的位置差;

(30) L_μ^R , μ 子计数器实际被击中的层数;

(31) L_μ^E , μ 子计数器预期被击中的层数;

除动量 p 之外, 其它量都或多或少地反映了三类粒子之间的差异. 它们都已适当地进行归一, 作为网络输入层的输入. 试验表明中间层的网点数在 15 到 20 之间变化不会明显改变网络的操作性能. 输出层 3 个网点, 分别对应于 e, μ, π 事例.

用于训练网络的 e, μ, π 样本事例数分别大于 5000 个, 以保证网络能学习到足够的粒子信息, 它们的动量谱如图 1 所示. 训练时将此三类事例依次提交给网络, 每次在每一类事例中均匀随机地选取一个事例作为输入, 设计输出层的第 i 个网点的输出目标值对应第 i 类事例为 1, 对应另外两类事例为 0. 网络的层间连结模式和训练算法基本同文献 [5], 只是将 S 型激活函数

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

中的温度参数 T 取为 0.4, 以使其表现出较强的阈函数行为, 提高压缩能力. 训练强度参数 η 和动量阻尼参数 α 在训练开始时分别取为 0.05 和 0.3, η 随训练次数的增加按

$$\eta_{N+1} = \eta_N \cdot 0.99 \quad (2)$$

方式逐渐减小, 但不小于 0.0001, α 当训练次数足够多后可取 0.2. 网点间连结的权重因子 w_{ij} 和阈因子 t_j 在 $[-0.1, 0.1]$ 区间随机地取初始值, 其改变量 Δw_{ij} 和 Δt_j 在每十个事例被训练后更新一次. 训练循环一般取 2×10^6 次, 若训练结果不满意, 可以把这时得到的 w_{ij} 和 t_j 作为初始值, 重新调整 η 和 α 参数进行再次训练, 这样可以大大节约再次训练的时间.

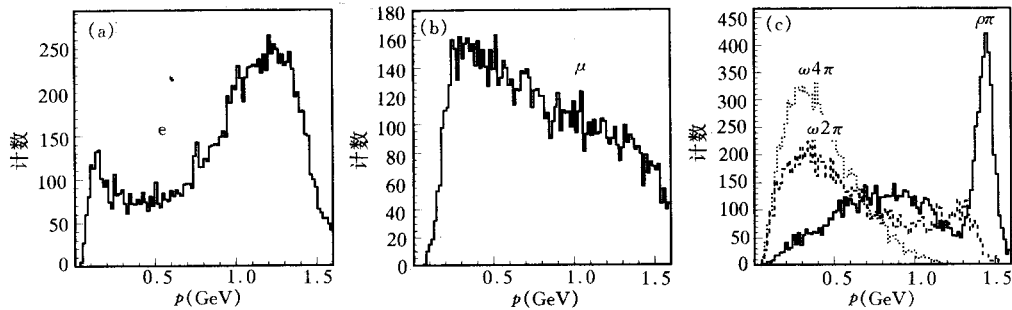


图 1 事例样本的动量谱

(a) 来自 $e e \gamma$ 道的 e , (b) 来自宇宙线的 μ , (c) 来自 $J/\psi \rightarrow \rho \pi, \omega \pi \pi, \omega 4 \pi$ 道的 π .

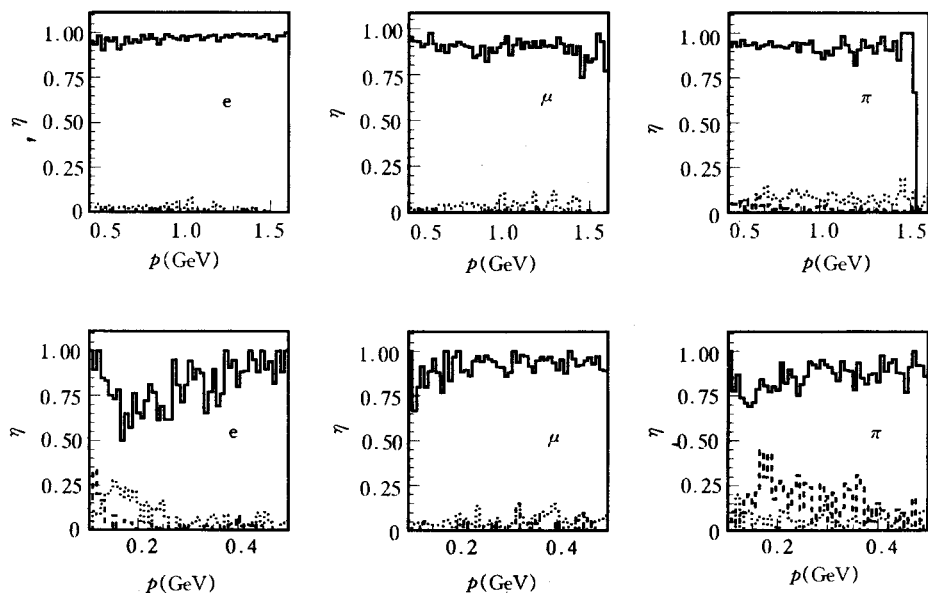


图2 BP网络设置各输出网点的截断值 $y_{\text{cut}} \geq 0.5$ 时给出的各类粒子的选择效率 η 和另外两类本底粒子的混入几率

(a) $p \geq 0.5 \text{ GeV}/c$, (b) $p < 0.5 \text{ GeV}/c$. 虚线和点线依次代表 μ, π 混入 e , e, π 混入 μ 以及 e, μ 混入 π 的几率.

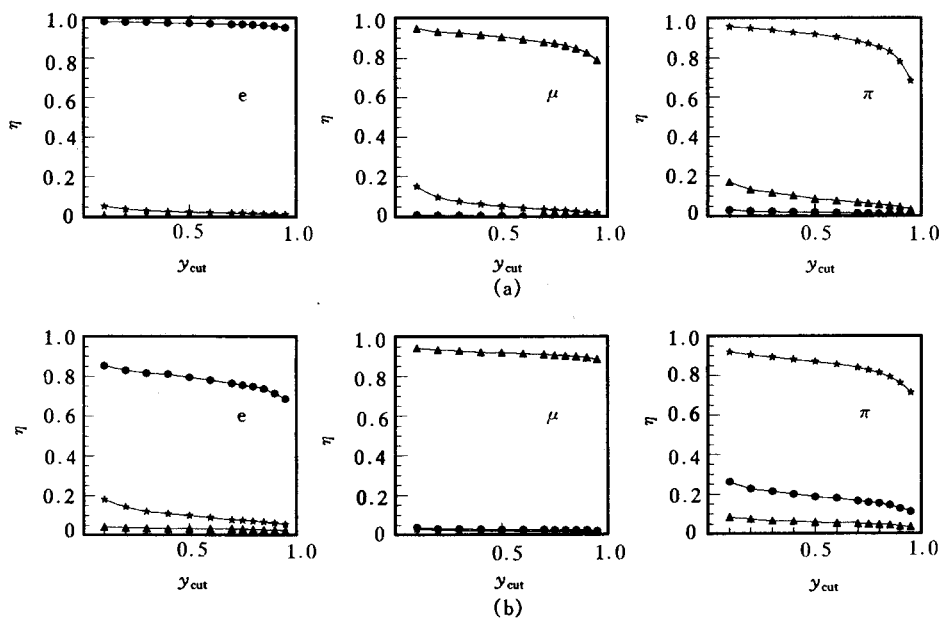


图3 各类粒子的选择效率 η 和本底水平随各输出网点截断值 y_{cut} 的变化

(a) $p \geq 0.5 \text{ GeV}/c$, (b) $p < 0.5 \text{ GeV}/c$. 图中 \circ 代表 e , \blacktriangle 代表 μ , \star 代表 π .

3.2 BP 网络的检验结果

BP 网络训练成功后, 固定所有的 w_{ij} 和 t_j , 将另一套独立的 e, μ, π 事例分别作为第一, 第二和第三类事例, 来检验网络的功能. 对各个输出网点设定不同的截断值, 就给出了各类事例的选择效率和来自另外两类事例的本底水平, 图 2 是设置各输出网点的截断值 $y_{cut} \geq 0.5$ 时给出的各类粒子的选择效率和本底粒子混入几率随动量变化的情况, 可以看出除了 $p=0.2\text{GeV}/c$ 附近的区域之外, 效率随动量的变化是相当均匀的. 若注意到图 1 所示样本动量谱的严重非均匀性, 这一结果是相当令人感兴趣的. 图 3 分别给出了各类粒子的选择效率和本底水平随各自输出网点截断值 y_{cut} 的变化.

为给出一个定量的概念, 在表 1 中列出了选取各输出网点的截断值为 $y_{cut} \geq 0.5$ 时 $p \geq 0.5\text{GeV}/c$ 的各类粒子的选择效率和另外两类粒子混入的几率, 同时给出了和 LVQ 网络结果的对比.

表 1 选取 BP 各输出网点截断值为 $y_{cut} \geq 0.5$ 时各类粒子的选择效率及与 LVQ 的比较

鉴别为	BP 网络输入			LVQ 网络输入		
	e	μ	π	e	μ	π
e	96.9%	0.3%	1.9%	87.6%	0.8%	1.8%
μ	0.4%	90.6%	5.3%	1.2%	87.1%	8.0%
π	1.9%	8.7%	91.9%	11.2%	12.1%	90.2%

4 LVQ 网络的训练和检验

4.1 LVQ 网络的结构和训练方法

表征某一粒子的 n 个物理量可视为 n 维特征空间的一个矢量, 用以描写该粒子. 不同的粒子事例所对应的特征矢量在该空间中应有不同的分布. LVQ 网络的基本操作思想是, 如果对第 c 类事例能找到一组最好地反应其特征分布的 l 个参考矢量, $m_i^c (i=1, 2, \dots, l; c=e, \mu, \pi)$ 被称为网络点, 那么对一个未知的事例 s , 若它和某一参考矢量有最小的距离 $|s-m_i^c|$, 则认为该事例属于 c 类. 可见 LVQ 网络训练的目的就是如何找出这些参考矢量, 我们采用了如下的作法^[6].

(1) 参考矢量的初始化. 对 e, μ, π 三类粒子, 每类给定 150 个初始参考矢量 $m_i^c(0)$, 分别由各自的数据样本中读入.

(2) 对每类粒子的参考矢量设置训练计数器 P_i^c , 初始值为 0.

(3) 将用以训练的 e, μ, π 样本事例, 依次等同地提交给网络, 对每一个事例 s , 找出与它距离最近的参考矢量 m_i^c , 设 s 属于 c' 类粒子, 按下列方法更新 m_i^c :

$$\begin{cases} m_i^c(t+1) = m_i^c(t) + \alpha_1(t) \cdot [s - m_i^c(t)], & \text{若 } c=c' \\ m_i^c(t+1) = m_i^c(t) - \alpha_2(t) \cdot [s - m_i^c(t)], & \text{若 } c \neq c' \\ m_j^c(t+1) = m_j^c(t) & \text{对 } j \neq i \text{ 的所有别的网点} \end{cases}, \quad (3)$$

α_1 和 α_2 是两个训练参数. 选取其初始值 $\alpha_1(0)=0.1, \alpha_2(0)=0.05$, 随着训练次数的增加逐

渐减少,

$$\alpha_{1,2}(t+1) = \alpha_{1,2}(t) \cdot 0.9. \quad (4)$$

(4) 若某一网点很少被训练到, 或者过多地被不同类的粒子事例所训练, 那么在下一训练循环前除去该网点. 根据训练计数器经反复试验调整, 最后选用了如下的截断条件, 即满足下面任一关系式的第 i 个网点将被舍弃,

$$P_i^{\text{total}} = \sum_c P_i^c \leq P_i^{\text{cut}} = 2; \quad (5)$$

$$f_i^{c_i} = \frac{\sum_{c \neq c_i} P_i^c}{\sum_c P_i^c} \geq f_i^{c_i, \text{cut}} = \begin{cases} 0.2, & \text{对 } e \\ 0.5, & \text{对 } \mu, \\ 0.5, & \text{对 } \pi \end{cases} \quad (6)$$

这样做以后如果某类粒子的网点数变得太少, 则可从事例样本中重新读入新的随机网点, 而后进行新的循环.

4.2 LVQ 网络的检验结果

图 4 是用另外一套独立的事例样本对训练好的 LVQ 进行检验所得到的结果, 给出了各类信号粒子的选择效率和另外两类粒子被混入的几率随动量的变化. 与图 2 给出的 BP 网络的结果相比较, 可以看出 LVQ 的选择效率明显低于 BP 网络, 而且随动量的变化也不如 BP 网络均匀. 表 1 定量地展示了这两类网络之间的差别.

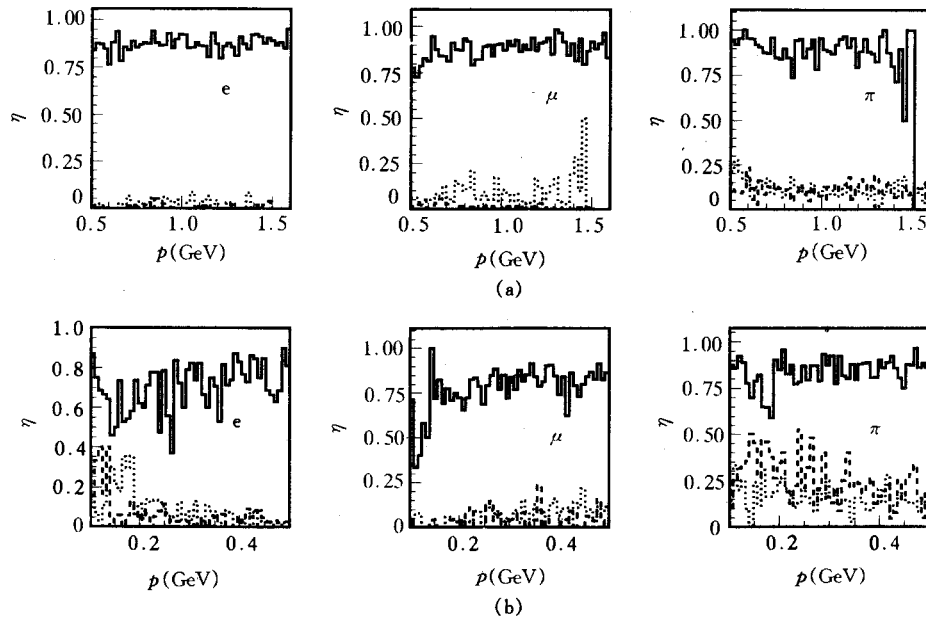


图 4 LVQ 网络给出的各类粒子的选择效率 η 和另外两类本底粒子的混入几率

(a) $p \geq 0.5 \text{ GeV}/c$, (b) $p < 0.5 \text{ GeV}/c$. 虚线和点线依次代表 μ 、 π 混入 e , e 、 π 混入 μ 以及 e 、 μ 混入 π 的几率.

5 结 论

本研究表明神经网络法可有效地应用于 BES 实验中的粒子鉴别, 即使在 M. C. 模拟不准确而不得不使用真实实验数据, 事例的分布严重不均匀的情况下, 网络仍然可以从样本学习到事例特征的详细信息. 这是一个非常令人感兴趣的结果, 具有重要的实用意义. 另一个结论是, 至少是在我们研究的这一课题中 LVQ 网络的模式识别功能不如 BP 网络强, 这同文献 [6] 所给出的结论是不同的. 该文献在 $\bar{p}p \rightarrow \bar{t}t$ 事例的鉴别中研究比较了几种方法, 认为 BP 网络法不如 LVQ, 而 LVQ 的结果又不如 Fisher 统计判别法, 但作者在文章中并没有给出利用 BP 网络法的详细鉴别结果.

参 考 文 献

- [1] Z. J. Jiang, T. J. Wang, Y. G. Xie *et al.*, *Nucl. Inst. and Meth.*, **A345** (1994) 541.
- [2] L. L. Ö nblad *et al.*, *Nucl. Phys.*, **B349** (1991) 675.
- [3] Wayne S. Babbage, Lee F. Thompson, *Nucl. Instr. and Meth.*, **A330** (1993) 482.
- [4] M. Joswig *et al.*, DESY 93-167, ISSN 0418-9833.
- [5] 张子平、王贻芳, V. Innocente, 高能物理与核物理, **18** (1994) 769.
- [6] A. Cherubini, R. Odorico, *Z. Phys.*, **C53** (1992) 139.

Identification of e, μ, π by Neural Network in BES

Zhang Ziping Chen Hongfang Ye Shuwei Zhao Jiawei

(Department of Modern Physics, University of Science and Technology of China, Hefei 230026)

Received 18 March 1996

Abstract

Two different kinds of neural network methods, BP and LVQ, are applied to the identification of e, μ, π particles in BES experiment, and better results are obtained. The μ data samples used for training and tests are from cosmic events, e and π are strictly selected from real experimental data by very tight cuts. Although their momentum spectrums are non-uniform, interesting enough is that the selection efficiencies given by BP test results are quite uniform in the whole momentum range. The results from LVQ is little worse. It shows that BP is more powerful in pattern recognition than LVQ at least in this study.

Key words BES experiment, particle identification, neural network method.