

高能物理实验中的前馈式神经网络法应用

张子平

(中国科技大学近代物理系)

人工神经网络是受生物神经网络的结构及其对外来信息处理方式的启示而得名的。它的研究起始于40年代,但由于当初提出的单层感知器模型固有的困难在长时间内困惑着人们,使得这方面的研究工作沉寂了几十年,直至80年代末,随着多层神经网络的提出,以及它在实际应用中显示的强大功能逐渐被人们普遍认识和接受,才重新刺激起这方面理论和应用研究的热情。

神经网络以其优于传统方法的强有力模式识别功能及抗噪声和本底干扰的强健性,近年来越来越受到高能实验物理学家的青睐,获得了广泛的研究和应用,在诸如粒子和衰变模式鉴别、共振态粒子不变质量的重建、高能带电粒子在探测器中飞行径迹的重建以及用于探测器在线数据获取系统二级触发的可能性等方面,都取得了很大的成功。

前馈式多层神经网络是一种最为成功和简单的人工神经网络结构。象一个一无所知的婴儿一样,为使其具有某种认知本领,实现某种期望的函数操作功能,则必需首先要用真的,正确的事例对其进行训练,使网络学习到足够的信息,这种训练和学习的方式称之为有监督的方式。别的结构的网络,如豪普费尔德网络,则可以是无监督的训练。本文在简要给出前馈式神经网络的基本结构和它的向后传播训练方法的基础上,着重介绍它在粒子鉴别、衰变模式和事例识别,以及不变质量重建方面一些成功应用的例子。

1. 前馈式网络的基本构架和训练

前馈式神经网络的基本构架可由图1表示,具有多层结构。第一层称为输入层,最后一层称为输出

层,中间可以有一个或多个中间层,或称为隐藏层。每层含若干个格点,称为网点或神经元。输入层的网点对应于事例的特征输入物理变量。后一层的各网点分别通过赋以权重值 W_{ij} 和阈值 t_i 的连线与前一层的各网点的输出相联系,例如由输入层到第一隐藏层,依图1的符号记法,首先求加权总和 $a_j = \sum W_{kj} x_k + t_j$,然后将一个激活函数(或称转移函数) $g(x)$ 作用到 a_j 上,就给出第一隐藏层第 j 个网点的输出为 $h_j = g(a_j)$, $g(x)$ 依具体问题可取为阶函数, S 形函数 $g(x) = 1/(1 + \exp\{-x/T\})$ 或 $\tanh(x/T)$, 或者它们的变型。 T 称为温度参数,不同的 T 值将改变 S 形函数的形状,控制网络的压缩能力, T 很小时接近于阶函数行

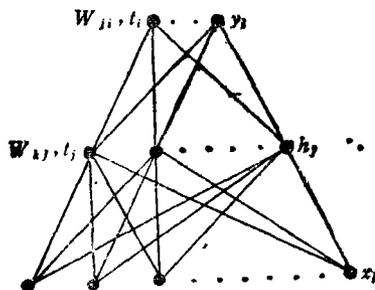


图1 前馈式神经网络的基本构架

为。 S 形函数将输出 h_j 压缩在 $[0, 1]$ 或 $[-1, 1]$ 区间,作为下一层的输入。其后各层的连接操作完全相同。

前面已经提到,这种前馈式网络一般采用有监督方式的向后传播法进行训练,基本原理和方法如下:

统取得更广阔的用武之地。二维实时光谱仪与太空条件的结合,将使太阳物理学取得具有重大历史意义的突破,这将包括磁场、大气动力学、射电毫秒级爆发,以及日地物理等我们具有优势的领域。另外,我们在2年多的空间太阳望远镜预研究中,对口径达1米的主光路系统,作了创新的设计,将以前的格里高利系统,改为反射系统。这不仅大大简化了结构,而且能消除附加偏振,从而实现“矢量场”的观测。这比美国和日本的系统有了重大改进。

一个口径1米的主望远镜系统,加上16通道的二

维光谱仪,构成了总重量达2.5吨的卫星的核心。这个卫星将运行在高度为500千米的太阳同步极轨的全日照轨道上。我国的长征4号B火箭就能完全满足发射要求。

“到太空去!”这是我国天文学家梦寐以求的理想;然而,一步登天毕竟也是不容易的。我们在技术上尚有一段磨合的过程。因此,第一步可以先进行气球中间实验,以促进条件的成熟。我期待在下世纪初,我国太阳物理能乘我国火箭卫星技术之东风,以实现该领域的历史性突破。

记由输出层第 i 网点的最后输出为 Y_i , 则训练网络实质上就是调整和最后选取最佳的权重值和阈值, 使得下面的误差函数取极小值。

$$E = 0.5 \cdot \sum_p \sum_i [Y_i(P) - A_i(P)]^2$$

这里的 P 指事例样本, i 是输出层网点号, A_i 是对输出设计的在 $[0, 1]$ 或 $[-1, 1]$ 区间的目标值。首先对所有的权重值和阈值在某一区间内随机地赋以初值, 该区间一般取 $[-0.1, 0.1]$, 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) 由输出层到最后一个隐藏层, 为使误差函数 E 逐渐减小, W_{ji} 和 t_i 应沿 E 的负梯度方向变化, 即它们的变化量 Δw_{ji} 和 Δt_i 应等于一个负常数 $-\eta$ 乘以误差函数 E 对 W_{ji} 或 t_i 的微分, 经过简单的运算可以得到 $\Delta w_{ji} = -\eta \delta_i g'(a_i) h_j + \alpha \Delta w_{ji}^d$, $\Delta t_i = -\eta \delta_i g'(a_i) + \alpha \Delta t_i^d$, 其中 $\delta_i = (Y_i - A_i)$, α 项称为动量项, 是为了抑制振荡而加入的, Δw_{ji}^d 指上一次循环得到的值, η 称为学习强度。

2) 对前面各层的操作原理和方法与上述相似, 如由最后隐藏层到其前一层, 通过简单的微分运算可知, 这时只需将上面式子中的 δ_i 换成 $\delta_i' = \sum \delta_i g'(a_i) w_{ji}$ 即可。

在训练中 α 一般在 $[0.2, 0.8]$ 之间试验选取, η 的取值区间为 $[0.001, 0.1]$, 开始时可取大些, 随着训练循环次数的增加逐渐减小。实践表明变化量 ΔW 和 Δt 在大致每 10 次输入后更新一次效果较佳。用于训练的正确事例样本最好要足够的大, 使网络得以充分地学到足够多的物理信息, 训练循环一般需要几万到百万次。

网络一旦训练成功后, 所有的权重因子和阈值因子都固定下来, 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。为反映网络的真实操作性能, 应使用另外一套完全统计独立的相同的事例样本对其进行检验。至此若向网络输入它以前未曾见过的事例, 则网络将会通过它在被训练时学到的, 存储于它的权重值和阈值中的知识来对新的事例进行分类和识别, 就象一个学会了区分苹果和梨子的孩童能正确地识别它们一样。

2. 粒子和衰变模式鉴别

Lund 大学 C. Peterson 等人建立了一个神经网络程序包 JETNET, 并且在 29 GeV 和 92 GeV 质心系能量下, 基于 JETSET 7.2、HERWIG 3.4 和 ARIADNE 3.1 模拟程序包进行的 e^+e^- 对撞模拟, 研究了夸克和胶子喷注、重夸克 (c 和 b) 喷注和轻夸克喷注的鉴别, 以及区分不同的强子化机制。采用两

种不同的输入变量方式, 对上述三种模型, 用一只只有一个输出网点的网络, 设置输出截断值 0.5, 鉴别夸克和胶子喷注都达到了 85% 的精度, 比贝耶斯理论极限只低 2% 左右, 而常规方法只有 65%, 最多也只能达到 70-75%。48-10-3 结构 (表示网络有三层, 第一层 48 个网点, 对应于事例的 48 个输入物理量, 中间层 10 个网点, 输出层 3 个网点, 下同。)的网络用以鉴别胶子、轻夸克和重夸克 (c 和 b) 喷注, 也得到了满意的结果, 区分 b 夸克喷注的效率和纯度可和顶点探测器相比较, 且对模型不灵敏, 表 1 是 $\sqrt{s} = 92 \text{ GeV}$ 能量下用不同模型的样本进行训练, 用 JETSET 样本检验得到的效率和纯度。

表 1 JETSET 样本对不同模型训练结果检验给出的效率和纯度 (括号内)

训练样本	胶子喷注 (%)	轻夸克喷注 (%)	重夸克喷注 (%)
ARIADNE	85(82)	53(72)	51(43)
JETSET	85(80)	50(72)	54(38)
HERWIG	84(79)	49(66)	47(34)

L. Bellantoni 等人用一个 25-5-1 结构的网络鉴别 LEP 能量下 $e^+e^- \rightarrow Z^0 \rightarrow q\bar{q}$ 过程中的 b 夸克喷注, 对经过 LEP e^+e^- 对撞机上的 ALEPH 探测器模拟的两喷注事例, 效率达 61%, 轻夸克喷注被误判为 b 夸克喷注的几率为 20%, 图 2 是输出网点给出的 b 夸克强子喷注和非 b 夸克强子喷注的事例分布。

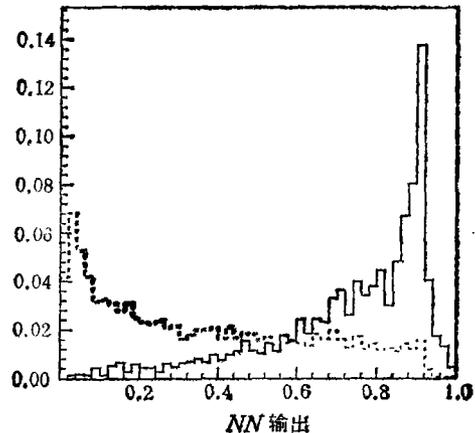


图 2 b (实线) 和非 b (虚线) 夸克强子喷注事例的网络输出分布

神经网络法已开始被应用于真实的实验数据物理分析中。ALEPH 组用一四层的网络研究了该探测器的 $\gamma-\pi^0$ 分辨, 证明网络法的分辨效率在 γ 或 π^0 能量的低能端和经仔细完善的常规方法相当, 而在高能端网络法的效率明显地高于常规方法, 图 3 是归一化

到97%的光子效率得到的 π^0 选择效率的网络法和常规方法比较。

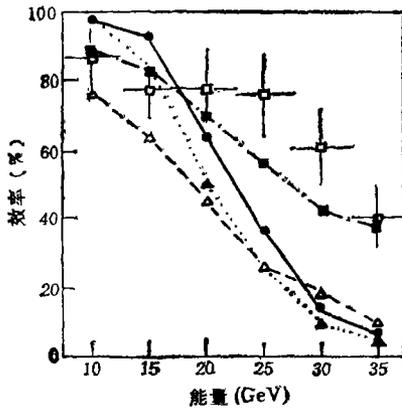


图3 网络法和常规方法得到的 π^0 选择效率(归一化到97%的光子效率),实和空方点分别对应于网络法的蒙特卡罗模拟和实验数据

ARGUS 实验组将网络法应用于 B 和 τ 物理的数据分析中,采用20-20-1结构,设置输出截断值 $Y_{cut} \geq 0.5$,从轻夸克对的连续本底中分辨 $e^+e^- \rightarrow \Upsilon(4s) \rightarrow B\bar{B}$,效率达 $(84.6 \pm 0.5)\%$,本底排斥率 $(80 \pm 0.6)\%$ 。用15-15-1结构的网络选择 τ 的半轻子衰变事例 $e^+e^- \rightarrow \tau^+\tau^- \rightarrow e\mu x$,设置输出截断值 $Y_{cut} \geq 0.985$,得到选择效率 $(24.7 \pm 0.3)\%$,本底9%,常规方法为效率 $(14.7 \pm 0.3)\%$,本底2%,最后分析给出 $B_c \cdot B_u = 0.0295 \pm 0.0007 \pm 0.0015$ 。另一部分工作是通过 τ 对的相关强子衰变道 $e^+e^- \rightarrow \tau^+\tau^- \rightarrow \pi^+\pi^0\nu_e\pi^-\pi^0\nu_e$ 末态带电 π 介子的角分布定矢量和轴矢量耦合的相干参数。定义规一化的矢量和轴矢量带电流耦合常数之积 r_{AV} ,设计4-10-4-1结构的网络,四个输入网点对应于 r_{AV}^2 等于0或1的四个特征角分布物理量,训练网络区分这两种分布。同时仔细地模拟本底,使实验数据和本底都通过训练好的网络产生输出,最后用网络输出的 r_{AV}^2 等于0或1的角分布谱拟合从数据减去本底而得到的实验分布,给出 $|r_{AV}| = 0.96 \pm 0.05 \pm 0.04$ 。

表2 在给定的输出截断值 Y_{cut} 下,训练(括号外)和检验(括号内)给出的各道的选择效率和本底水平

衰变道	Y_{cut}	选择效率(%)	本底水平(%)
$\tau \rightarrow \rho\nu$	0.8	54.8(53.1)	15.9(18.1)
$\tau \rightarrow \pi\nu$	0.9	65.6(63.2)	14.6(15.6)
$\tau \rightarrow e\nu\nu$	0.9	75.9(76.1)	1.8(1.9)

下面介绍一下我们在L3的 τ 物理分析中用网络法鉴别 $\tau \rightarrow \rho\nu_e$ 、 $\tau \rightarrow \pi\nu_e$ 和 $\tau \rightarrow e\nu_e\nu_e$ 的工作。在 $\tau \rightarrow \rho\nu$ 的极化分析中,由于洛伦兹效应,L3探测器不能

很有效地区分 π^0 和单个 γ ,为此我们设计了一个18-18-1结构的网络用以事例的筛选,表2是在给定的输出截断值 Y_{cut} 下各道的选择效率和本底水平,表3是对 $\tau \rightarrow \pi\nu_e$ 和 $\tau \rightarrow e\nu_e\nu_e$ 道与常规方法的比较,以证明网络法的正确性和实用性。用于 $\tau \rightarrow \rho\nu$ 道的极化分析,得到 $p_\tau = -0.129 \pm 0.050 \pm 0.050$ 。

表3 网络法与常规法选择效率和本底水平的比较

衰变道	$\tau \rightarrow \pi\nu$		$\tau \rightarrow e\nu\nu$	
	常规筛选	网络筛选	常规筛选	网络筛选
选择效率(%)	57	63.2	74.5	76.1
本底水平(%)	8	15.6	4.2	1.9
数据事例数	1156	1515	3202	3239
相同数据事例	812		3036	
蒙特卡罗事例数			5009	5027
相同蒙特卡罗事例			4806	

3. 不变质量的重建

共振态粒子不变质量的重建是高能物理实验数据

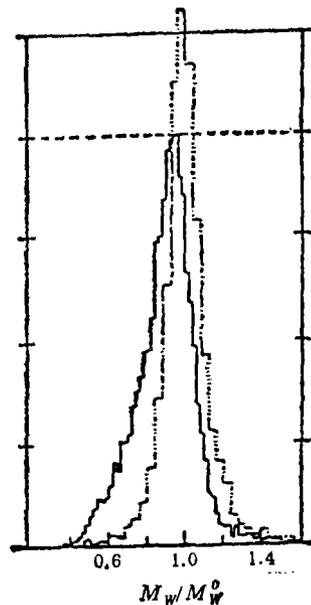


图4 常规法(实线)和网络法(虚线)重建的质量(M_w)除以真正质量(M_w^0)的分布

分析中重要的课题,例如在 $p\bar{p}$ 对撞中 $W \rightarrow q\bar{q} \rightarrow$ 强子的过程,W的质量由末态强子的动量和能量重建,麻烦在于末态强子有部分是来自P或 \bar{P} 的剩余部分子,

新一代高能 γ 射线望远镜系统

王 焕 玉

在高能天文观测领域,实现对硬 X 和 γ 射线源的巡天搜索;高精度定位以至成像,始终是引人密切关注的一个重要方向.本文简单介绍编码孔径成像原理和技术,并介绍一种新的实际的旋转编码孔径成像 γ 射线望远镜系统.

硬 X 射线(20—120keV)和 γ 射线天文学的发展,需要高灵敏、高精度的望远镜系统.由于硬 X 和 γ 射线能量高于 3keV 时,会聚光学开始变得无能为力.从而造成了在这一领域内制作成像望远镜的特殊困难.

近年来,从事高能天文观测的科学工作者想了很多办法,希望能解决这一日益迫切需要解决的难题.日本宇航所小田等人提出用多层或多个不同间隔的条栅型旋转调制准直器实现成像.于 1978 年同 UCSD 合作在气球上观测蟹状星云,得到了 22—63keV 能区的强度分布图像.角分辨为 20 弧秒.但是,随着能量的提高,准直器制作困难,用这种方法实现对硬 X、 γ 射线的观测就有一定困难.1981 年秋,丹麦空间所曾用板条镶嵌的双闪烁探测器和旋转调制准直器,对 γ 爆炸作了定位精度约 10 弧分的观测.而近几年采用最多的是编码孔径成像技术.这一技术是在多针孔成像、随机条形孔径方法和富里埃变换谱仪等技术上发展起

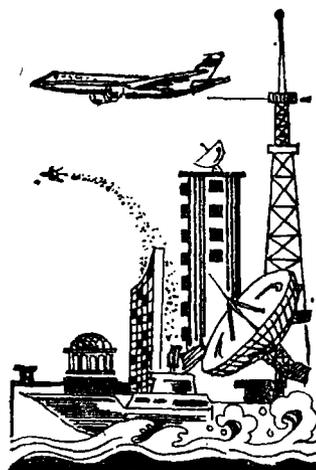
来的.这种成像方法,不需要任何镜头,使探测器能够同时观测源的全部区域.这种方法只需要一个栅网和位置灵敏探测器,不要求精密机构,因而发展较快.

它的特点是可实现宽视场观测,延伸源成像、点源的高精度定位.原则上可用于无针孔衍射效应的软 X 射线到几十 MeV 高能 γ 射线的任意能区.

编码孔径 γ 射线望远镜的基本思想是由针孔成像发展而来.因此,在描述编码孔径 γ 射线望远镜的原理之前,先简单介绍一下针孔成像及其功能.

① 针孔成像

在一块硬纸板上有一个小针孔,把它对着光源(例如图 1 所示),在纸板的另一侧放一块毛玻璃,适当调节它的位置,在毛玻璃上可以获得光源的清晰的倒像,这个现象,就是我们十分熟悉的小孔成像.显然,针孔



且 q 或 \bar{q} 还会辐射出胶子喷注,因此鉴别合适的喷注是常规法重建 W 不变质量的关键.基于蒙特卡罗模拟,一个四层的前馈式神经网络被设计用于 W 质量的重建,结果优于常规法,如图 4 所示,这是因为网络法更好地捕捉住了辐射胶子喷注的信息.

质量重建中一类很重要的问题是许多共振态粒子在其衰变末态中,由于有中微子而产生大的丢失能量,因而用常规的方法往往无法重建它的不变质量,我们最近的研究表明网络法可以在这方面大有作为.例如对 pp 对撞中的黑格斯粒子衰变道 $H^0 \rightarrow \tau^+ \tau^- \rightarrow e \mu \nu$ 事例,特别是高亮度下有堆积效应(pile-up)存在时,用网络法都正确地重建了 H^0 的质量,这对 LHC pp 对撞机上高亮度下通过 $H^0 \rightarrow \tau^+ \tau^-$ 道寻找中能质量区的 H^0 粒子是非常重要的.

4. 前馈式神经网络的性能和优缺点

(1) 具有很强的模式识别功能,及很好的抗噪声和本底干扰的强健性(Robustness).

(2) 和传统的判别式(Discriminant)方法相比,它可以非常直接地近似和实现任意的连续函数操作功能,无需对各个物理量及截断条件进行非常仔细的研

究、构造和选取.做为这一优越性能的反面,一般情况下人们不能对网络内部的操作功能给出解析的描述.

(3) 网络结构的并行方式适宜于计算机并行运算,以及 VLSI 可编程电子学硬件的实现,因此将一改传统的诺曼(Von Neumann)串行计算机模式,实现高速度的函数操作功能,使其应用于探测器在线数据获取系统二级触发成为可能.

(4) 由于前馈式神经网络采用有监督的向后传播法进行训练,其函数操作功能依赖于训练样本的精确性,如蒙特卡罗模拟的精度.传统的用于事例筛选的判别法也是如此.许多研究说明,网络的这一误差是可以进行分析估计的,而由网络自身引入的误差一般并不大,并且也是可以控制和分析的.

(5) 局域极小值问题.若误差函数有多个极小值,在训练过程中则有可能陷入局域极小,而不能实现理想的操作功能.但在一般情况下,函数都是足够光滑的,并且通过开始时选取较大的训练强度 η 值,而后随着训练次数的增加逐渐减小,大多可以避免陷入局域极小.现在也提出了许多解决局域极小值问题的方法,如成功有效的玻尔兹曼机制等.