NEURAL NETWORK: A POWERFUL TOOL FOR CLASSIFICATION

YIJIN PEI*
I. Physikalisches Institut, RWTH Aachen
D-52056 Aachen, Germany

Artificial Neural Networks have been widely used in the data analysis of the L3 experiment at LEP. Different appoaches are applied on various physics topics such as photon identification, tau decays and b-tagging. Encouraging results have been obtained. Systematic errors due to the use of neural networks have been extensively studied.

1. Introduction

It has been shown that neural networks (NNs) are very efficient for feature classification, pattern recognition and combinatorial optimization.[1,2] In general, to deal with tasks where the classification problem is not linearly separable, the input variables have to be mapped into an intermediate space, where the probability distributions are linearly separable. This can be usually achieved by a multilayer feedforward NN,[1,2] where the hidden layers perform approximately the desired mapping of the input variables. Another approach is to use a functional-link NN,[2] where the input variables are first expanded into a space, which is spanned by a basis of orthonomal functions. In this space the classification problem is linearly separable.

In this paper we show two examples of the application of NNs in the L3 data analysis. In the first one we describe a functional-link NN, which is used to identify photons in hadronic Z decays. In the second one we show how to classify hadronic one-prong tau decays by using feedforward NNs, with emphasis on the evaluation of the systematic errors due to the use of the NNs.

2. Identification of Photons in Hadronic Z Decays

The L3 electromagnetic calorimeter,[3] composed of about 11000 bismuth germanium oxide (BGO) crystals, has excellent energy and spatial resolution for photons over a wide energy range. The conventional way to select photons[4] requires strong isolation cuts on the photon cluster, which leads to a low selection efficiency for photons inside jets. By using NNs, the correlations between the variables that characterize the photon shower can be exploited. For this reason a better separation between the photon signal and the background can be achieved.

2.1. *The functional-link neural network*

In this approach each input variable x_k is first scaled to be in the range $(-1, 1)$. It

*present address: CERN/PPE, CH-1211 Geneva 23, Switzerland.

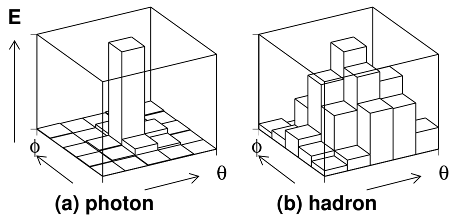

Fig. 1. Shower produced by (a) a photon and (b) a hadron in the BGO calorimeter.

is then expanded using orthonormal functions

$$f(x_k) = w_{1k}x_k + \sum_{j=1}^{n}(w_{2kj}\sin(j\pi x_k) + w_{3kj}\cos(j\pi x_k)) \tag{1}$$

where w is a set of weights, n the order of expansion. Usually $n = 4$ is enough for our problems. The transformed variables now serve as input to a classifier, which is a single Perceptron node. A Sigmoid function is used as the activation function:

$$F_c(X) = \frac{1}{1+e^{-X/T}}, \quad \text{with } X = \sum_{k=1}^{N} f(x_k) + w_0. \tag{2}$$

Electromagnetic and non-electromagnetic showers in the BGO calorimeter have different shape, as shown in Fig.1. We choose variables which are most sensitive to the shower shape as inputs. They include the energy E_9 measured in 9 crystals centered on the most energetic crystal, the fraction of energy deposited in each of those 9 central crystals, the energy measured in the 9 central crystals divided by that measured in the 25 central crystals (E_9/E_{25}), and a χ^2 like variable, χ^2_{BGO} which compares the fraction of energy deposited in each of the 9 central crystals with a typical electromagnetic one (calculated from test beam data).

For training the NN we use the Monte Carlo(MC) events generated by the programs JETSET7.3,[5] with a full simulation of the L3 detector response.[6] The weights w are determined by minimizing $[F_c(X) - F_{\mathrm{known}}]^2$, where F_{known} is chosen to be 0.975 for the photon and 0.025 for the background. Since the shower shape of photons and hadrons in the BGO calorimeter varies with the energy, the NN is trained separately in four energy ranges: 0.1–1, 1–3, 3–5 and 5–45 GeV. Each training set comprises 10,000 to 100,000 patterns depending on the energy range.

2.2. *Performance of the neural network*

To test the NN we use another set of MC events (test sample) which are independent of those used for the training. The NN output F_c for BGO clusters in all energy ranges is shown in Fig.2(a). We see a good separation between the photon signal and the background. Also shown in Fig.2(a) is the NN output for the data. There is a slight difference between the data and MC distribution. Since the NN is very sensitive to the details of the shower shape, small differences between data and MC

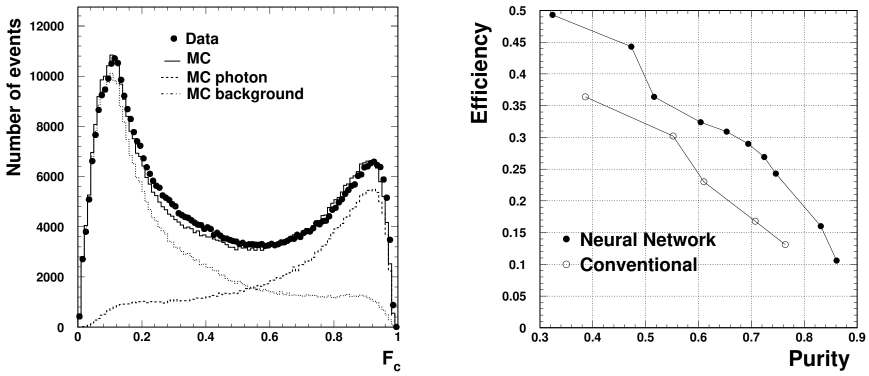

Fig. 2. (a) Distribution of the NN output F_c for data and MC; (b) Efficiency-purity plot for photons in all energy ranges.

in the input variables are amplified in the NN output. In Section 3.2 we show an example how to evaluate this kind of systematics.

The performance of the NN is compared to that of the conventional L3 photon selection. In the case of the NN selection we use $F_c > 0.5$, while in the conventional selection we require $\chi^2_{BGO} < 5$ and $E_9/E_{25} > 0.8$. Fig.2(b) shows the efficiency-purity plot for photons in all energy ranges. This plot is obtained by changing the cut values on F_c, χ^2_{BGO} and E_9/E_{25}. As one can see from the plot, the NN approach improves the efficiency by about 40% for purities between 60% and 70%, which are used in most of the analyses. For neutral mesons which decay into two or more photons, such as π°, η and ω, the selection efficiency is improved by more than a factor of two.

3. Classification of Hadronic One-Prong Tau Decays

In this analysis, we use feedforward-backpropagation NNs to classify hadronic one-prong τ decays into a single charged particle, h, plus $n\ \pi^\circ$ ($\mathrm{h} = \pi^\pm, \mathrm{K}^\pm$; $0 \le n \le 3$).

3.1. *Selection of hadronic one-prong tau decays*

The selection of hadronic one-prong tau decays is described in detail in Ref.[7]. Because of the large Lorentz boost at LEP energies, photons from the decay chain $\tau \to \mathrm{h}\ \mathrm{n}\pi^\circ, \pi^\circ \to \gamma\gamma$ are significantly merged together in the BGO calorimeter. Fig.3(a) shows the distribution of the number of neutral clusters in τ decays with contributions of the various sources indicated. For 1, 2, 3 and 4 reconstructed neutral clusters, we use 3, 3, 2 and 1 NNs, respectively, to classify τ decays into the four categories ($\mathrm{h}, \mathrm{h}\ \pi^\circ, \mathrm{h}\ 2\pi^\circ, \mathrm{h}\ 3\pi^\circ$). The NNs have 17 to 22 input variables depending on the number of reconstructed neutral clusters. The input variables include energy of the neutral clusters, momentum of the charged track, variables which describe the shower shape, and all angles between the neutral clusters and the track as well as the inter-cluster angles. All NNs have only one hidden layer, which has the same number of nodes as the input layer. Each output layer has only

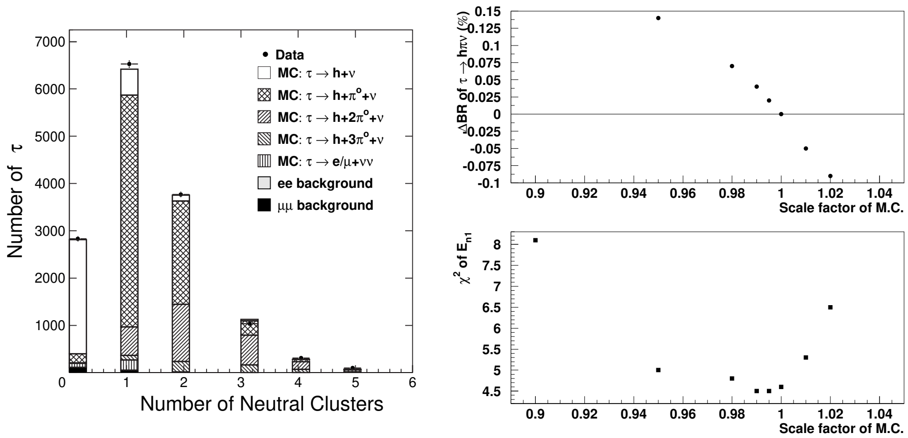

Fig. 3. (a) Distribution of the number of neutral clusters in τ decays; (b) Change in BR($\tau \to$ h $\pi^{\circ}\nu$) and χ^2 caused by scaling the energy of the first neutral cluster in the MC.

one node. A Sigmoid function is used as the activation function.

The training sample of 250 000 $\tau^+\tau^-$ events is generated with KORALZ3.8,[8] with a full simulation of the L3 detector response. A test sample (independent of the training sample) of 110 000 $\tau^+\tau^-$ events is used to determine the selection efficiency for a τ decay which is classified into one of the four categories. The efficiencies, which include both the efficiencies in the $\tau^+\tau^-$ event selection and the efficiencies in the NN selection, are shown in Table 1. The probabilities to accept an event from the two main background sources are also shown there.

3.2. *Determination of branching fractions*

From a data sample corresponding to an integrated luminosity of 21.68 pb^{-1} we select 2967, 6613,1060 and 293 decays in the four decay categories (h, h π°, h $2\pi^{\circ}$, h $3\pi^{\circ}$), respectively. We determine branching fractions using the relation

$$N_i^{\exp} = N_\tau \sum_{j=1,10} \mathrm{BR_j}\epsilon_\mathrm{j}^\mathrm{i} + \sum_{\mathrm{k}=1,2} \mathrm{N_k^{bg}}\epsilon_\mathrm{k}^\mathrm{i}, \tag{3}$$

where $N_i^{\exp}$ is the number of events expected in category i, N_τ the total number of τ decays, N_k^{bg} the number of background events, and ϵ_j^i the selection efficiency for a τ decay channel j which is classified into category i as specified in Table 1.

For the branching fractions into h $4\pi^{\circ}$, h η π°, $\pi\mathrm{K_L^0}$, $l\nu\nu$ and three-prong we use the results given in Ref.[9,10,11]. The branching fractions $\mathrm{BR_j}$ ($1 \leq j \leq 4$) are determined using a χ^2 fit which compares the predicted number and the observed number of decays in the four categories. We obtain:

$$\begin{aligned}
\mathrm{BR}(\tau \to \pi/\mathrm{K}\ \nu) &= (11.82 \pm 0.26(stat) \pm 0.43(sys))\% \\
\mathrm{BR}(\tau \to \pi/\mathrm{K}\ \pi^{\circ}\ \nu) &= (25.05 \pm 0.35(stat) \pm 0.50(sys))\% \\
\mathrm{BR}(\tau \to \pi/\mathrm{K}\ 2\pi^{\circ}\ \nu) &= (8.88 \pm 0.37(stat) \pm 0.42(sys))\% \\
\mathrm{BR}(\tau \to \pi/\mathrm{K}\ 3\pi^{\circ}\ \nu) &= (1.70 \pm 0.24(stat) \pm 0.38(sys))\%
\end{aligned}$$

Table 1. Efficiencies to select a given decay channel in each of the four categories.

Index in Eq.(3)	Decay Channel	Selection Efficiency (%)			
		h	h π°	h $2\pi^\circ$	h $3\pi^\circ$
$j = 1$	π/K ν	56.56	3.33	0.30	0.03
2	π/K π° ν	2.50	65.09	1.23	0.15
3	π/K $2\pi^\circ$ ν	0.24	9.73	26.14	3.78
4	π/K $3\pi^\circ$ ν	0.09	2.09	9.22	21.28
5	π/K $4\pi^\circ$ ν	–	0.71	3.00	20.20
6	π/K $\eta\pi^\circ$ ν	–	1.20	3.40	15.60
7	$\pi K^0_L \nu$	27.36	17.72	0.84	0.18
8	e ν ν	0.98	0.79	0.21	–
9	μ ν ν	0.80	0.61	0.02	–
10	3-prong	0.08	0.21	0.06	–
$k = 1$	e^+e^-	0.03	0.004	–	–
2	$\mu^+\mu^-$	0.26	0.04	–	–

Contributions to the systematic error for each decay channel are listed in Table 2. The largest contribution comes from the systematic uncertainty due to the NN selection, which is caused mainly by the discrepancy between data and MC in the NN input variables and in the NN output. The contribution from the NN output is estimated by varying the cut on each NN's output by 10%. This contribution becomes larger as the separation between signal and background decreases. See, for example, the π/K $2\pi^\circ$ ν and π/K $3\pi^\circ$ ν channels in Table 1 and 2.

Because of the large number of the input variables used, it is a formidable task to evaluate the contribution to the systematic error from each input variable. However, since most of the input variables are correlated, we only have to study uncorrelated variables or variables with large influence on the NNs' decisions. The latter are identified by applying a shift, a scale or smearing factor to the inputs one by one for the data following the discussion in Ref.[12]. By observing the systematic changes in the extracted branching fractions, we find that the momentum of the track, electromagnetic and hadronic energies and the transverse electromagnetic energy, as well as the energies and angles of the three most energetic neutral clusters have the largest influence on the result. We then shift, scale or smear those variables in the MC and calculate the χ^2 of the fit of the MC distribution of those variables to that of the data. We calculate also the change in the efficiencies and thereafter the change in the branching fractions (see Fig.3(b)). The change corresponding to a $\Delta\chi^2$ of one from the optimum shift, scale or smearing is taken as the size of the contribution to the systematic error.

We have also studied input variables which were not selected for the above studies. We find that their influence on the result is indeed very small. Furthermore we have performed checks like changing the parameters in the network training, changing the MC distribution by removing a fraction of events at random to best fit the data distribution, and testing the stability of the networks using the methods described in Ref.[12]. Their contributions to the systematic error are small.

As an independent cross-check, a separate analysis was performed using cuts in the quantities with most separation power between channels. The results are consistent with the values obtained by the neural network method.

Table 2. Contributions to the systematic error (in %) on the branching fractions.

Source	ΔBR(π/Kν)	ΔBR(π/K$\pi^\circ\nu$)	ΔBR(π/K$2\pi^\circ\nu$)	ΔBR(π/K$3\pi^\circ\nu$)
$\tau\tau$ selection	0.14	0.20	0.11	0.17
N_τ	0.10	0.22	0.08	0.02
MC stat.	0.18	0.26	0.18	0.15
E_n threshold	0.03	0.07	0.07	0.10
h selection	0.16	–	–	–
Form factor	–	–	0.18	0.09
ΔBR(h $4\pi^\circ$)	–	–	0.02	0.06
ΔBR(h η π°)	–	–	0.01	0.02
ΔBR(πK$^0_L\nu$)	0.08	0.03	0.01	–
NN output	0.04	0.13	0.22	0.18
NN input	0.29	0.28	0.22	0.20
Total	0.43	0.50	0.42	0.38

4. Conclusions

We have applied NNs on various physics topics in the L3 experiment. The performance of the NNs is much better than that of the conventional methods. We have shown that a reasonable estimation of the systematic errors due to the use of NNs can be achieved.

Acknowledgements

I would like to thank T. Aziz, G. Chen, G.M. Chen, V. Innocente and Y.F. Wang for many helpful discussions. This work was supported by the German Bundesministerium für Forschung und Technologie.

References

1. R. Hecht-Nielsen, *Neurocomputing*, (Addison-Wesley Publishing Company 1990).
   B. Denby, "Tutorial on Neural Network Applications in High Energy Physics: A 1992 Perspective", Fermilab Preprint, FERMILAB-Conf-92/121-E.
   R. Beale and T. Jackson , *Neural Computing: An Introduction*, (Adam Hilger Publishing, Bristol 1990).
2. Y.H. Pao, *Adaptive Pattern Recognition and Neural Networks* (Addison-Wesley Publishing Company 1989).
3. L3 Collab., B. Adeva *et al.*, *NIM* **A 289** 35 (1990).
4. L3 Collab., B. Adeva *et al.*, *Phys. Lett.* **B 286** 403 (1992).
5. T. Sjöstrand and M. Bengtsson, *Comput. Phys. Commun.* **43** 367 (1987);
   T. Sjöstrand, *Z Physics at LEP*, CERN report CERN-89-08, Vol.III, p.143. (1089)
6. The L3 detector simulation is based on GEANT3.14. See R. Brun *et al.*, "GEANT 3", CERN DD/EE/84-1 (Revised), September 1987. The GHEISHA program (H. Fesefeldt, RWTH Aachen Preprint PITHA 85/02 (1985)) is used to simulate hadronic interactions.
7. M. Acciarri *et al.*, L3 Collab., CERN preprint CERN-PPE/94-178; November 11, 1994.
8. S. Jadach, B.F.L. Ward and Z. Was, *Comp. Phys. Comm.* **66** 276 (1991).
9. Particle Data Group, L. Montanet *et al.*, *Phys. Rev.* **D 50** (1994).
10. J. Urheim, "Second Workshop on Tau Lepton Physics", p. 165, edited by K.K. Gan, World Scientific, 1993; T. Skwarnicki, same proceedings, p. 217.
11. CLEO Collab., M. Artuso *et al.*, CLNS 94/1281.
    HRS Collab., R. Tschirhart, *et al.*, *Phys. Lett.* **B 205** 407 (1988).
12. V. Innocente, Y.F. Wang and Z.P. Zhang, *NIM* **A 323** 647 (1992).