ISTITUTO NAZIONALE DI FISICA NUCLEARE

Sezione di Trieste

INFN/AE-92/14

27 aprile 1992

S. Bianchin, G. Cosmo, P. Garlatti Costa, A. De Angelis, M. DeNardi, B. Denby, M. Dickson, P. Eerola, J. Kalkkinen, L. Lanceri, A. Linussio, G. Pauletta, L. Santi, C. Vignaduzzo, A. Zoppolato and N. Weiner

THREE APPLICATIONS OF NEURAL NETWORKS FOR HIGH ENERGY PHYSICS

Servizio Documentazione
dei Laboratori Nazionali di Frascati

Three Applications of Neural Networks for High Energy Physics

S. Bianchin, G. Cosmo, P. Garlatti Costa, A. De Angelis, M. DeNardi, B. Denby, M. Dickson, P. Eerola, J. Kalkkinen, L. Lanceri, A. Linussio, G. Pauletta, L. Santi, C. Vignaduzzo, A. Zoppolato and N. Weiner

Abstract

This note collects three presentations to the Second International Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, held in Lalonde sur Maures (France), in January 1992. The three presentations deal with classification problems: events in Z^0 decays, pions in an electromagnetic calorimeter, and jets in hadronic collisions.

DELPHI RESULTS ON THE MEASUREMENT OF THE PARTIAL HADRONIC WIDTHS OF THE Z^0 USING NEURAL NETWORKS

G. COSMO, A. DE ANGELIS and A. LINUSSIO
Istituto di Fisica dell'Universita' di Udine and INFN Trieste
Via Fagagna 208, I-33100 Udine, Italy

P. EEROLA and J. KALKKINEN
Research Institute for High Energy Physics
Siltavuorenpenger 20 C, SF-00170 Helsinki 17

ABSTRACT

The measurement of the hadronic branching fractions of the Z^0 boson into all five known quarks, from the data collected by the DELPHI detector at LEP during 1990, is presented. The measurement was based on four single-output neural networks.

Progress is reported on the design of a more powerful network architecture, and on the calculation of systematic errors.

1. Introduction

DELPHI has measured for the first time in the literature [1] the partial hadronic widths into the five flavours kinematically available from the decay of the Z^0 boson.

The measurement was performed by coupling four single-output feed-forward neural networks, specialised respectively in the classification of $b\bar{b}$, $c\bar{c}$, $s\bar{s}$, and $(u\bar{u}+d\bar{d})$ unresolved, and then by using the rate of final-state radiation to estimate the relative production of quarks of charge 2/3 to quarks of charge 1/3 [2]. The work in Ref. [1] was extending previous studies aiming at the classification of decays into $b\bar{b}$ pairs via neural networks [3].

The preliminary results from the neural network in Ref. [1] were

$$\Gamma_{u\bar{u}+d\bar{d}}/\Gamma_h = 0.417 \pm 0.015\,(stat) \pm 0.058\,(sys) \quad (1)$$

$$\Gamma_{s\bar{s}}/\Gamma_h = 0.233 \pm 0.016\,(stat) \pm 0.051\,(sys) \quad (2)$$

$$\Gamma_{c\bar{c}}/\Gamma_h = 0.139 \pm 0.010\,(stat) \pm 0.058\,(sys) \quad (3)$$

$$\Gamma_{b\bar{b}}/\Gamma_h = 0.211 \pm 0.006\,(stat) \pm 0.020\,(sys)\,, \quad (4)$$

and two main problems were left to be solved:

1. Correlations between the four networks. The branching fractions were extracted by constructing an overall χ^2 from the four networks, keeping into account at first order the covariances. The smallness of the statistical error with respect to the systematic one made us confident that keeping into account covariances at all orders should not change much the result. Nevertheless, the problem could be solved by constructing a single network with four output nodes, that maps the n input variables into a space R^4, in which the four classes could be separated via multidimensional analysis.

2. Correct estimate of systematics. Main source of systematics is related to model-dependence of the classification. The simulated events used for the classification of the real ones depend on some physical parameters known with errors. Of course a complete treatment of systematics should include the effect of variation of such parameters, inside the limits in which they are bounded when comparing shape variables to Monte Carlo predictions, *without any hypothesis or assumption on the branching fractions.* Such confidence intervals are not provided by the literature at present, and an effort to calculate them is in progress.

A work has started in DELPHI to overcome these two problems. In this note we summarise the results obtained with the four networks technique, and the work that has been done related to point 1, that is at present almost solved.

This analysis is based on a data sample collected by the DELPHI detector [4] at LEP during 1990, with an experimental selection procedure described in Ref. [5].

2. Four Networks Technique

Eighteen variables were used as an input for the separation.

Their choice came from the examination of the literature, and from a study of flavour-dependent distributions based on the JETSET 7.2 Parton Shower Monte Carlo [6] (JETSET PS in the following), that has proven, after two years of activity of LEP, to reproduce well the main features of the hadronic decays of the Z^0 [5,7]. The particles in an event were clustered in jets according to the JADE/E0 algorithm [8], with $y_{cut} = 0.05$. In the following, the most energetic jet will be called "first jet", and indicated by the superscript (f); the second most energetic jet will be called "second jet", and indicated by the superscript (s).

Description of the variables follows.

1. The sphericity $S^{(f)}$ of the first jet, calculated after a boost $\beta = 0.96$ along its axis. The axis of the jet was defined by the sum of the momenta of the particles belonging to it.

2. The directed sphericity $S^{(f)}_{1234}$ of the 4 most energetic particles in the first jet. For a set Q of tracks in a jet, this variable is defined as

$$S_Q = \frac{\sum_Q p_t^2}{\sum_Q |p|^2}$$

where the $|p|$'s are the momenta in the rest frame of the set Q and the p_t's are their components perpendicular to the original jet direction in the laboratory frame.

3. The directed sphericity $S^{(s)}_{1234}$.
4. The invariant mass $M^{(f)}_{1234}$ of the 4 most energetic particles in the first jet.
5. The invariant mass $M^{(s)}_{1234}$ of the 4 most energetic particles in the second jet.

6..9. The products of the homologue direct sphericities for triplets of particles in the first and the second jet, $S^{(f)}_{ijk} \times S^{(s)}_{ijk}$.

10..13. The products of the homologue invariant masses for triplets of particles in the first and the second jet, $M^{(f)}_{ijk} \times M^{(s)}_{ijk}$.

14. The momentum of the slowest pion of the jet 1, after a boost along the jet axis corresponding to a D^* energy equal to one half of the beam energy.
15. Same as 14., for the second jet.
16. The momentum of the most energetic K^0 in the event (0 if no kaons reconstructed).
17. The momentum component perpendicular to the axis of the nearest jet of the most energetic K^0 in the event (0 if no kaons reconstructed).
18. The sum over the jets of the ratios between the momentum of the leading particle and the momentum of the jet.

All variables were rebinned in such a way that they were ranging from 0 to 1.

Four independent feed-forward neural networks were used (one for each class that was separated from the others) with 18 nodes in the input layer, associated with the input variables x_i, defining the pattern space P; a variable number of nodes in the hidden layer; and one output node, associated with the output value Θ, belonging to the feature space F. We refer to [1] for details on the training of the networks, based on back-propagation, and on their architecture.

In the training phase, the four networks were specialised in such a way that network "1" was designed to be more performant for separating Z^0 decays into $u\overline{u}$ or $d\overline{d}$, network "2" for separating decays into $s\overline{s}$ pairs, network "3" for separating decays into $c\overline{c}$ pairs, and network "4" for separating decays into $b\overline{b}$ pairs.

From each of the four networks ($i = 1...4$), the fraction of events $\beta_j^{(i)}$ of each class j ($j = 1...4$) was determined by means of a χ^2 fit to the form

$$\Re^{(i)}(t) = \sum_j \beta_j^{(i)} a_j^{(i)}(t)\,,$$

where $\Re^{(i)}(t)$ is the map of the data through the network i into the feature space, and $a_j^{(i)}(t)$'s are the distributions for each class j in the feature space, both determined in the test sample. All distributions were normalised to unity.

The four networks were constructed in such a way that each network provided a fit with small correlation coefficients between the class that the network itself was teached to distinguish and the other classes.

Finally, the expression

$$\chi^2 \simeq \sum_i < \vec{\beta}^{(i)} - \vec{\beta}^* | \mathcal{C}^{(i)-1} | \vec{\beta}^{(i)} - \vec{\beta}^* >\,,$$

where $\mathcal{C}^{(i)}$ is the covariance matrix in the fit from the i-th network, was minimised with respect to $\vec{\beta}^*$, under the constraint that the sum of the branching fractions is equal to 1. This led to the determinations and to the statistical errors quoted in (1..4).

The study of systematics kept into account the uncertainties in the best tuning of parameters in JETSET PS as parametrised in [7]. In addition, a detailed study of the effect of fragmentation parameters was done, to obtain the systematic errors in (1..4).

3. One Network with Four Output Nodes

Passing from a one-output neural network to a four output one represents a rather nontrivial qualitative step. The neural network should now perform a mapping $\mathcal{F} : P \subset R^{23} \longrightarrow F \subset R^4$, where P is the pattern space and F is the feature space.

The network was trained to separate four classes of events *simultaneously*: each class was associated to one output node and the flavour of the event was interpreted to be the one, the node of which gave the largest output value. For every input pattern, therefore, only one of the four output neurons was taken as output 1, all the others 0. The input pattern was described by 23 variables; 5 new variables were added to the previous set of 18. They are:

19. Sum of the track impact parameters, each one scaled by the error (to reinforce the classification of $b\bar{b}$ events). Tracks with impact parameters greater than 2 mm are omitted because they are likely to come from secondary decays of strange particles.

20. Charge flow ($Q_f - Q_b$). The event is divided in two hemispheres by the thrust axis, and the axis is oriented so that it is pointing always to the forward hemisphere (polar angle $< 90^0$). The charge in the forward (backward) hemisphere is calculated by weighting the charge of each track by its momentum and dividing the weighted sum of charges by the sum of the absolute values of the track momenta, i.e., $Q_f = (\sum_f |\vec{p}|Q)/(\sum_f |\vec{p}|)$, for all tracks in the forward hemisphere. This variable is sensitive to the quark charge, and the asymmetry of the quark polar angle (due to quark electroweak couplings).

21. Sum of the absolute values of the charges of the two most energetic jets, weighted by the momenta in the same way as in 19. This variable is sensitive to the quark charge.

22. Absolute momentum $|\vec{p}|$ of the most energetic lepton (0 if no leptons with momentum > 2 GeV, which is the identification limit).

23. $|\vec{p_t}|$ of the most energetic lepton with respect to the axis of the closest jet.

The network structure was completed by one hidden layer of 9 neurons, fully connected to the input and output layers.

The training phase of the network was performed on a set of 86,990 Z^0 decays into $q\bar{q}$ pairs. The training set was composed by an equal number of patterns for each flavour type. The entire training process was divided into a defined number of sessions (*training epochs*). During each of them, the training set was filtered by the network. The weights in the network were initialised with small random values in the range [-0.1,0.1] and were updated every 10 input patterns according to the generalised delta-rule.

After each training the *learning parameter* η and the *momentum parameter* α were geometrically updated, ranging from 0.05 to 0.0001 and from 0.4 to 0.9 respectively, while the *temperature* T was fixed ($T = 2.0$).

The testing phase of the network was performed using a different set of 73,691 simulated events, independent from the training set. The results presented in Table 1 were obtained after a test session performed on a training of 200 iterations. The quality of the separation of flavours is comparable with the one obtained with the four networks described in Section 2.

The results on the hadronic branching fractions of the Z^0 boson can be obtained by minimising a χ^2 measure of the discrepancy between the four flavour densities from simulated Monte Carlo data and the experimental one in a four-dimensional hypercube. The technique involves Multivariate Discriminant Analysis, using an optimisation algorithm to fit a discriminant surface between the events of one flavour and all the others in R^4 [10].

At present, we are working on a technique (Parzen windows, [11]) to smooth the probability density function for the simulated data in R^4, in order to reduce

Testing	Efficiency	Purity
$(u\overline{u}+d\overline{d})$	44.9 %	55.8 %
$s\overline{s}$	28.4 %	31.4 %
$c\overline{c}$	36.0 %	29.7 %
$b\overline{b}$	75.7 %	60.1 %

Table 1: *Network performance in Testing.*

the statistical fluctuations coming from finite sampling.

Results in Table 1 indicate the feasibility of the measurement.

References

1. C. Bortolotto et al., "A Measurement of the Partial Hadronic Widths of the Z^0 using Neural Networks", Udine Preprint 91/04/AA (August 1991), to be published in the Proceedings of the Workshop on Neural Networks, Isola d'Elba (Italy), June 1991.
2. A. De Angelis, "Hadronic Branching Fractions of the Z^0 boson", Udine Preprint 91/08/AA (November 1991), to be published in the Proceedings of the International Symposium on Multiparticle Dynamics, Wuhan (China), September 1991.
3. L. Lönnblad et al., Nucl. Phys. **B349** (1991) 675.
   C. Bortolotto, A. De Angelis and L. Lanceri, Nucl. Instr. and Meth. **A306** (1991) 459.
   L. Bellantoni et al., Nucl. Instr. and Meth. **A310** (1991) 618.
   P. Henrard (ALEPH Coll.), presented at the 4th Symposium on Heavy Flavour Physics, Orsay, June 1991.
4. P. Aarnio et al. (DELPHI Coll.), Nucl. Instr. and Meth. **A303** (1991) 233.
5. P. Aarnio et al. (DELPHI Coll.), Phys. Lett. **B240** (1990) 271.
6. T. Sjöstrand, Comp. Phys. Comm. **27** (1982) 243, ibid. **28** (1983) 229.
   T. Sjöstrand and M. Bengtsson, Comp. Phys. Comm. **43** (1987) 367.
7. M.Z. Akrawy et al. (OPAL Coll.), Z. Phys. **C47** (1990) 505.
8. W. Bartel et al. (JADE Coll.), Z. Phys. **C33** (1986) 23.
9. DELSIM User's Guide, DELPHI 89-67 PROG 142, Geneva, July 1989.
   DELSIM Reference Manual, DELPHI 89-68 PROG 143, Geneva, September 1989.
10. A. Linussio, Thesis at the University of Udine (1992).
11. E. Parzen, "Modern Probability Theory and its Applications", John Wiley and Sons, New York 1960.

A NEURAL NETWORK FOR e/π CLASSIFICATION IN A CALORIMETER

P. GARLATTI COSTA, A. DE ANGELIS, L. LANCERI,
L. SANTI, C. VIGNADUZZO and E. ZOPPOLATO
Istituto di Fisica dell'Universitá di Udine and INFN Trieste
Via Fagagna 208, I-33100 Udine, Italy

ABSTRACT

A neural network (NN) classifier is under study, for the separation of electromagnetic from hadronic showers in a highly-segmented Si/W calorimeter. The performance of the algorithm is studied on simulated data, for a detector inspired by the prototype of the calorimeter for the WIZARD experiment.

1. Introduction

The gross sketch of the detector, inspired by the prototype of the calorimeter for the WIZARD experiment[1], consists of a tower of XY sampling layers, alternated with showering material (W) planes. The sense layers are formed by two SI-D detectors, each divided in strips and mounted back to back to give X and Y coordinates of sampling.

Main features of detector are summarized in Table 1.

Number of sampling layers	20
Number of strips for each Si-D detector	16
Width of strips	3.6 mm
W planes thickness	$0.5X_0$
Total area	$(60 \times 60)mm^2$
$R_{Molière}$	0.9 cm

Table 1: - Main features of the detector.

The granularity of detector provides ability of measuring the shower profile for particle identification with high accuracy, also in the absence of full shower containment (e.g., for high energy electrons).

The simulation of this apparatus has been performed with GEANT 3.14 Monte Carlo program. The response of each sense strip (i.e., the energy deposed) was estimated by means of the track length of the secondary electrons and positrons generated in the shower.

2. The Input Variables

Although the detector was designed to generate XY samplings, we used for this study only the information coming from the X type strips. After the calibration, converting data into Minimum Ionizing Particle (MIP) units, and the noise suppression, the signals are elaborated to recognize clusters and to calculate the variables used for the event reconstruction.

The variables used by the NN are the following (see Appendix for more details):

I1) Total cluster dispersion normalized: the sum of widths of all clusters, scaled to total energy $\Sigma_{tot}/E_{tot} = \frac{\sum_k^{(clus)} \sigma_{cl_k}}{\sum_i e_i}$

I2) Total dispersion of clusters normalized: the sum of the dispersion of the centers of clusters inside each plane, scaled to total energy $\Delta_{tot}/E_{tot} = \frac{\sum_j^{(plane)} \delta_{pl_j}}{\sum_i e_i}$

I3) Internal/External Energy ratio: Ratio of the sum of the energies deposed on each plane in strips within $2 \times R_{Molière}$ from x_{pl_j} to the sum on the energies deposed outside the same range E_{int}/E_{ext}

I4) Longitudinal development: it measures the *thickness* of the shower, compared to a MIP $S_{lon} = \frac{\sum_j^{plane}(E_{pl_j} - MIP)}{\sum_i e_i}$

Fig. 1 shows the distributions of variables I1-I4 for showers of 2 GeV.

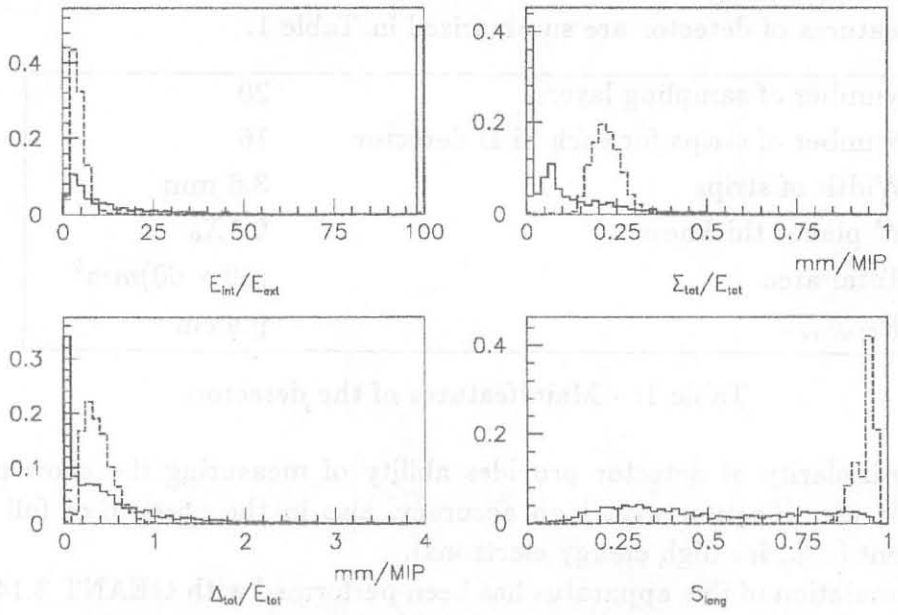

Figure 1: - Network input variables I1-I4 for showers of 2 GeV. Dotted lines represent data from e showers, solid lines from π showers.

3. The Network

The NN used for this study is a Feed Forward NN (FFNN) with 4 input nodes, 3 nodes within 1 hidden layer and 1 output node. Activation is given by means of a sigmoidal function $g : R \rightarrow (0., 1.)$, with temperature T .

Training was performed with the Gradient Back Propagation algorithm (GBP), based on a standard quadratic error function and with a momentum term. Learning was tuned by modifying any of the δ-learning rule parameters, that are allowed to vary exponentially during the training, from an initial value to the final one.

The update of weights had been performed every 10 events.

The training of NN was performed on a set of 1000 events randomly chosen with equal probability from the two classes of showers, and submited to the network over 5 epochs. The test phase was performed with a different set of data, of the same size and composition.

During test we used data containing showers for e of 2, 4, 6 GeV and π of 2, 4 GeV. Three different kinds of energy mixture were used: all data mixed up, 4 GeV e and 2 GeV π, 6 GeV e and 2 GeV π.

Table 2 summarizes the results obtained by applying two different sets of parameters during learning and thresholds for the discrimination.

Learning parameters and fiducial regions for NN output in the classification		e^- 2,4,6 GeV π 2,4 GeV		e^- 4 GeV π 2 GeV		e^- 2 GeV π 2 GeV	
		e^-	π	e^-	π	e^-	π
α=0.5 η=0.5 T=2.0	Efficiency	0.97	0.92	0.98	0.95	0.96	0.98
$O_{e^-} \in [0.0,0.5)$ $O_\pi \in [0.5,1.0]$	Purity	0.93	0.97	0.95	0.98	0.98	0.96
α=0.5-0.7 η=0.5-0.1 T=2.0-0.4	Efficiency	0.95	0.92	0.99	0.94	0.97	0.96
$O_{e^-} \in [0.0,0.1]$ $O_\pi \in [0.9,1.0]$	Purity	0.93	0.95	0.94	0.99	0.96	0.97

Table 2: - Discrimination efficiency and purity obtained with the FFNN.

Fig. 2 shows a scheme of the system used in the development and test of the network.

The training has been performed on a minicomputer μVAX 3000 and data (i.e., the weights of the NN) were downloaded to a transputer network, through the host file server (an IBM PS Personal Computer), to perform the tests.

The transputer network had been simulated by a single INMOS T800, with 4 Kbyte of static RAM, baud rate of 20 Mbit/s for the physical links, and 3.5 MFLOPS of peak performance[2].

After elaboration, results were transmitted to a WorkStation for analysis and representation.

The algorithm for the event reconstruction can be divided in two classes of

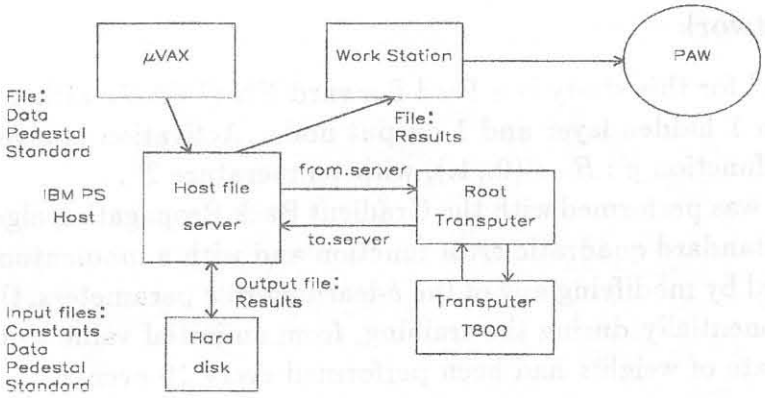

Figure 2: - The system setup and data flow used in the study.

subprocesses: local variable calculations (of quantities related to a plane of the calorimeter) and global variable calculations. Table 3 contains the results of timing tests performed for each class of subprocesses and the total time consumed by the different tasks.

The synchronization of data transmission takes the largest part of the time. By supposing that elaboration of local information can be performed in parallel, we estimate the elaboration rate to be 140 event/s.

	Local variables (μs)			Global variables (μs)		
	Mean	Max	Min	Mean	Max	Min
e	423	704	192	1408	1472	1408
π	495	768	192	1408	1472	1408

Synchronization time (for 20 planes)	5054 μs
Elaboration time for planes variables (parallel)	768 μs
Elaboration time for global variables (serial)	1408 μs
Elaboration time for FFNN discrimination	139 μs
Total elapsed time	7367 μs

Table 3: - Results of timing test of the reconstruction and classification algorithm.

4. The Unsupervised NN Approach to the Problem of e/π Classification

Besides the study of the FFNN performances in e/π classification above described, we have tested the application of a NN with Kohonen[3] structure to the problem.

We used an array of 20 $\times$ 20 nodes, connected planarly. The response of each node of the network was given by the euclidean distance between the vector of the input data and the weight vector associated to the neuron.

The response of network was defined as the node with the minimum distance. In the training phase the weights $\omega_i^{(j)}$ of the j^{th} node inside a zone of *neighbourhood* around the winning node were modified, following the learning rule

$\Delta\omega_i^{(j)} = \eta(t)(x_i(t) - \omega_i^{(j)}(t))$ with $\eta(t) = A_\eta e^{-\lambda t}$.

The width β of the neighbourhood zone during the training was allowed to decrease with time, to resolve progressively narrower structures: $\beta(t) = A_\beta + B_\beta e^{-\lambda' t}$.

Typical values of the parameters used were:

$A_\eta = 0.8$ $A_\beta = 1.$ $B_\beta = 8.$ $\lambda = 2/((\#$ of data) $\lambda' = 1.15\lambda$

The network has been trained to discriminate two different kinds of set of variables: one related to the longitudinal shower profile and the other to the lateral shower profile. The profiles are defined as the number of activated strips, projected in the first case on the beam axis and in the second case on a plane perpendicular to that axis.

Tests have been performed on 2 GeV data. Fig. 2 shows an example of the results of test.

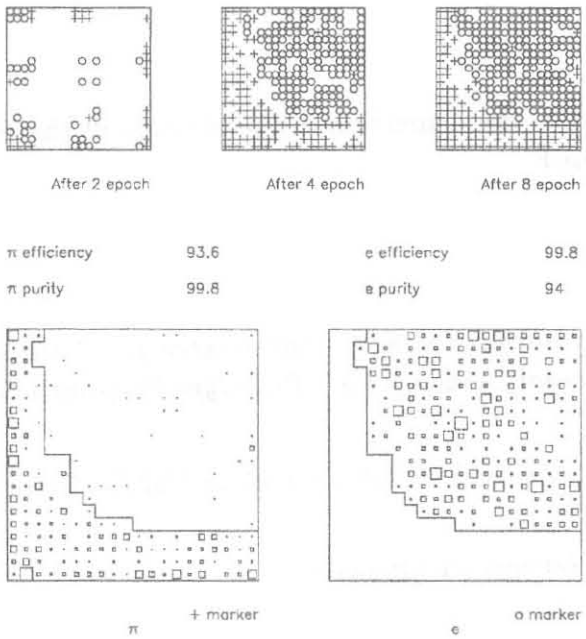

Figure 3: - Kohonen NN response to lateral profile data, 2 GeV.

The upper half of the figure shows three stages of evolution of the network trained with the lateral profile data. In each diagram a circle represents a node which registered more e than π and a cross a node which registered more π than e. The activation status of the nodes evolves until almost all nodes are classified.

The lower part of Fig. 3 shows the final state of the network in more detail. The size of the boxes is proportional to the number of activations of a node, for π on the left and for e on the right. After training, efficiency and purity have been calculated from the number of events inside each fiducial regions for π and for e, delimited in the figure by the polyline.

Appendix: The Definition of the Variables

Given a set S of strips, we define the variables

Energy in Set	$E_S = \sum_{i \in S} e_i$
Baricenter of Set	$x_S = \frac{\sum_{i \in S} e_i x_i}{E_S}$
Dispersion in Set	$\sigma_S = \sqrt{\frac{\sum_{i \in S} e_i (x_i - x_S)^2}{E_S}}$

were x_i is the position of strip i^{th}.

By means of these quantities the following variables are defined:

For the k^{th} cluster (set of contiguous firing strips in a sense plane)

$E_{cl_k}, x_{cl_k}, \sigma_{cl_k}$

for the j^{th} plane

$E_{pl_j}, x_{pl_j}, \sigma_{pl_j}$

and a variable related to the dispersion of the clusters in the planes

$$\delta_{pl_j} = \sqrt{\frac{\sum_{cl_k \in pl_j} E_{cl_k} (x_{cl_k} - x_{pl_j})^2}{E_{pl_k}}}$$

References

1. G. Barbiellini et al., *Silicon Tracking Calorimeter for Antimatter Search in Space: First Experimental Results from a Prototype* Preprint LNF-91/048 (IR), 9 July 1991.
2. M. Homewood et al. *The IMS T800 Transputer* IEEE Micro, October 1990, V.7 N.5, p. 10.
3. T. Kohonen, *Neural Networks* **1** (1988) 1.

CLASSIFICATION OF JETS FROM $p\bar{p}$ COLLISIONS AT TEVATRON ENERGIES †

S.BIANCHIN[1], DENARDI[1], B. DENBY[2], M. DICKSON[3],
G. PAULETTA[1], L. SANTI[1] and N. WAINER[2]

[1] *Istituto di Fisica dell'Universita' di Udine and INFN Trieste*
Via Fagagna 208, I-33100 Udine, Italy
[2] *Fermi National Accelerator Laboratory, Batavia Ill. USA*
[3] *Dept. of Physics, Univ. of Rochester*

ABSTRACT

The use of feed-forward neural networks for the identification of jets from $p\bar{p}$ collisions has been investigated. Simulated data was used to train and optimize the network and it's effect on real data was evaluated qualitatively. An alternative method of discrimination was also investigated for comparison

1. Introduction

The identification of the parton ancestors of jets has drawn a lot of attention in the course of the last decade. Many prescriptions[1], based both on theoretical[2] and empirical[3] criteria have been proposed. Nevertheless, uncertainties in the hadronization models remain a major obstacle in the detailed testing of QCD.

The problem is twofold; (i) identifying the jet characteristics or "variables" most sensitive to the differences between quark and gluon jets and (ii) developing the most efficient discriminant function. Neural networks (NNs) are well-suited to solving the latter aspect of the problem and may also contribute to the former. It is therefore not surprising that NNs have recently been applied to it's solution, the more so because the might be applied on-line. Most of the work[4,5] has involved the application of feed-forward networks to discriminate between jets produced in e^+e^-collisions at LEP energies. However, some work[6] has also been done for jets from $p\bar{p}$ collisions at CERN energies.

In this paper, an investigation into the possibility of applying feed-forward NNs to discriminate between quark and gluon jets at CDF (Collider Detector at Fermilab) is presented. The network was optimized and trained using simulated data generated by means of standard Monte Carlo's (MCs)[7] and "smeared" by propagating it through a full detector simulation. The choice of variables was guided by conventional wisdom[1,2,3] and by experience at CDF. They are: (i) the charge multiplicity; (ii) the sum of the transverse momentum p_t for tracks inside a cone of radius $r \leq 0.2$ around the jet axis, where $r = \sqrt{\eta^2 + \phi^2}$, η is the pseudorapidity and ϕ is the azimuthal angle; the analogous sums for (iii) $0.2 \leq r < 0.4$ and; (iv)

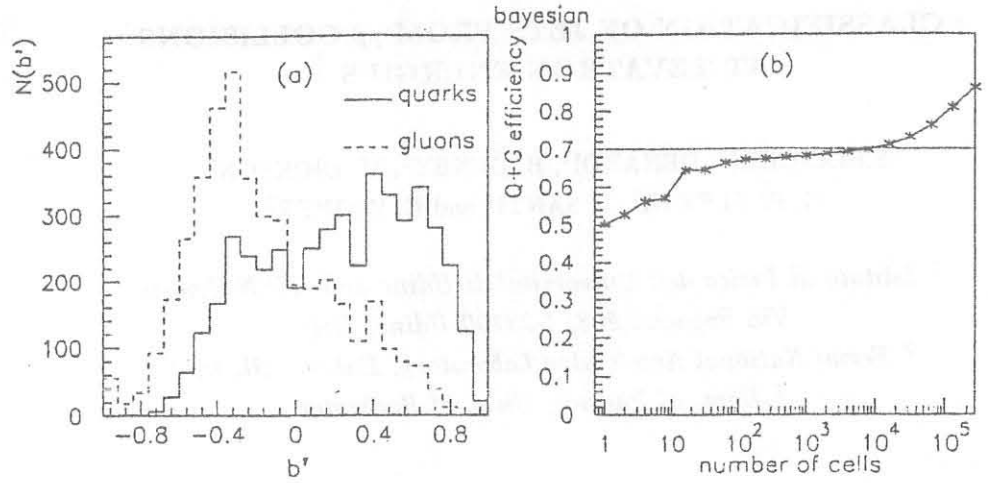

Figure 1: *Bayes classification of PYTHIA jets; (a) quark and gluon b'-distributions; (b) the average identification efficiency ϵ as defined in the text*

$0.4 \leq r < 0.7$; the variances of the transverse energy (E_t)-weighted distributions of (v) η and; (vi) ϕ ;(vii) p_t (w.r.t the jet axis) of the leading particle in the jet and; (viii) the distance (in $\eta - \phi$) space between the jet axis and the leading particle. Note that the p_t sums are normalized to E_t and that E_t is not used explicitly as a variable because, as opposed to dijets from e^+e^-collisions, the total cm. energy is unknown.

Bayes classification (section 2) was used to evaluate the discriminating power of these variables. A feed-forward NN was then optimized using MC-data (section 3) and applied (section 4) to different MCs to evaluate model-dependence. Flavor-dependence was studied by tagging MC quarks and the NN's effectiveness with real data was evaluated qualitatively. An alternative discriminant (Fisher discriminant) was also applied for comparison (section 5).

2.BAYES Classification

A Bayes classifier was constructed and applied to $N = 1000$ events consisting of equal numbers $N_q = \alpha_q N$ of quarks and $N_g = \alpha_g N$ of gluons ($\alpha_q = \alpha_g = 0.5$). The distributions of quarks and gluons with respect to the modified Bayes probability $b' = 2b - 1$, where the Bayes probability $b = n_q/(n_q + n_g)$ and n_q and n_g are the numbers of quarks and gluons in a cell of feature-space, is shown in fig. 1a. The average classification efficiency $\epsilon = (Nq(> thresh) + N_g(< thresh))/N$ for a b'-threshold of 0, is shown in fig. 1b. A classification efficiency of about 70% is attainable with a signal-to-background ratio of 3:1. The rise on the right is due to limited statistics.

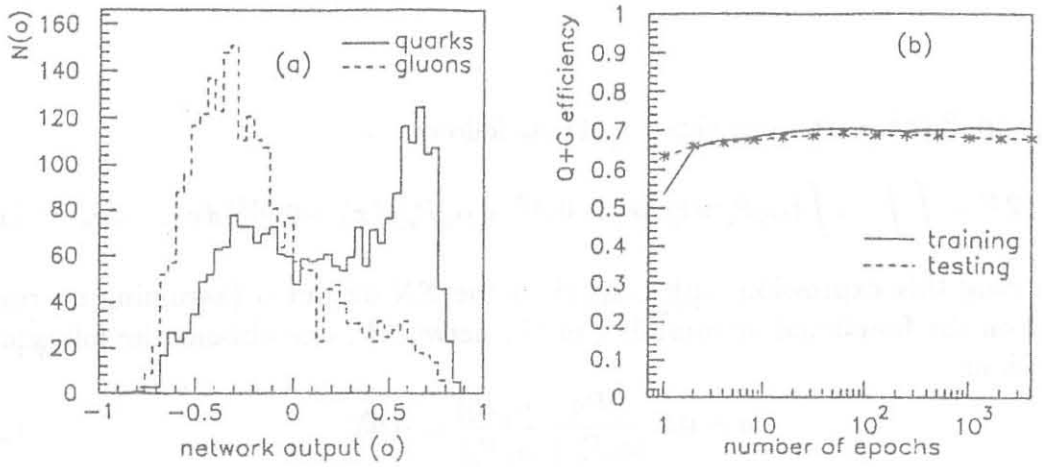

Figure 2: *Results of propagating the "test" data set from PYTHIA through the NN trained with the same MC; (a) output distributions and; (b) the average identification efficiency as defined for the Bayes classifier (see text). Also indicated (solid line) is the efficiency for the training set.*

3. The feed-forward network

The software simulation of a feed-forward NN containing 8 inputs, one hidden layer with 6 neurons and a single output was trained, using Gradient Back Propagation (GBP), on simulated data; a set of 5000 simulated jets containing equal proportions of quarks and gluons. A similar but independent set was used to test generalization. The network was optimized for our application. Criteria for the selection of the network architecture may be summarized as follows; (i) suggestions[8] that a "symmetric" sigmoid (e. g. one that varies between -1 and $+1$, rather than 0 and $+1$) speeds up convergence of GBP were substantiated; (ii) a simple perceptron was observed to obtain all but a few percent of the best attainable classification efficiency but; (iii) a few more neurons were needed to guarantee the functional flexibility required to render the output distribution independent of the starting point for training (see also the discussion below in connection with the Bayes classifier); (iv) too many neurons were seen to lead to overspecialization and; (v) more than one hidden layer contributes only to increasing the convergence time of GBP.

The NN was trained to associate output values of -0.9 and $+0.9$ with gluon and quark jets, respectively. The result of 4096 passes through the entire training set ("epochs") is shown in fig. 2. The best attainable efficiency is around 70% and the generalizing power of the network is seen to be good. A comparison of the output distributions shown in fig. 4a with the analogous Bayes distributions shown in fig. 1a reveals a striking similarity. This similarity is not casual as has already been remarked in[9] . For a NN trained on a set of N jets containing quarks and gluons in proportions α_q and α_g, respectively ($\alpha_q + \alpha_g = 1$), the error E, minimized by GBP, may be expressed in terms of the probability distributions

$P_q(x)$ and $P_g(x)$ (each normalized to 1), as follows;

$$2E = \int\int \ldots \int \{\alpha_q P_q(x)[o(x) - 0.9]^2 + \alpha_g P_g[o(x) + 0.9]^2\} dx_1 \ldots dx_n \qquad (1)$$

Minimizing this expression with respect to the NN output o (assuming no constraint on the functional adaptability of the network[9] , one obtains the following expression;

$$o = 0.9\frac{[\alpha_q Pq - \alpha_g P_g]}{[\alpha_q P_q + \alpha_g P_g]} = 0.9b' \qquad (2)$$

The similarity between the distributions with respect to b' and o is therefore explained; the NN approximates the Bayes classifier and, to the degree that it succeeds, obtains a discriminant function that can not be improved upon. The degree to which it succeeds depends on it's functional adaptability; i.e. on the number of neurons in the hidden layer. These considerations have had a bearing on our choice of network architecture as mentioned above.

4. Application

If an arbitrary set of N' jets containing α'_q and α'_g ($\alpha'_q + \alpha'_g = 1$) proportions of quarks and gluons is propagated through the NN, the resulting output distribution $P'(o)$ will be given by;

$$P'(o) = \alpha'_q P_q(o) + (1 - \alpha'_q) P_g(o) \qquad (3)$$

Eq. 3 may obviously be used to extract unknown values of α'_q, assuming the distributions P_q(o) and $P_g(o)$ can be determined reliably from training on simulated data. One can fit these distributions to the data to extract α_q *real* jets. The success of this method depends on how well the MC simulation represents reality and this is not easy to gauge. An idea of the model-dependence of this method may be had by using an NN trained on one MC to distinguish between quark and gluon jets generated by another. The MC HERWIG[7] was used alternately with PYTHIA[7] for this purpose. Some model-dependence (about 15%) was observed. Fig. 3a shows the results of using a PYTHIA-trained network to determine α_q from data generated by HERWIG. Flavor-dependence was also investigated by propagating "tagged" jets through the NN. From fig. 3b., one could conclude that B quarks are gluon - like. Further investigation revealed that this similarity decreases with increasing energy and that B quarks regain their quark - like characteristics above 30 GeV.

A preliminary evaluation of the applicability of MC - trained NNs to real data was also made. A detailed analysis will appear in a later work. Fig. 4 shows the result of propagating jet-20 (events taken with a 20 GeV trigger threshold on the jet E_t) data from CDF and dijet events from the AMY detector[10]. It appears that

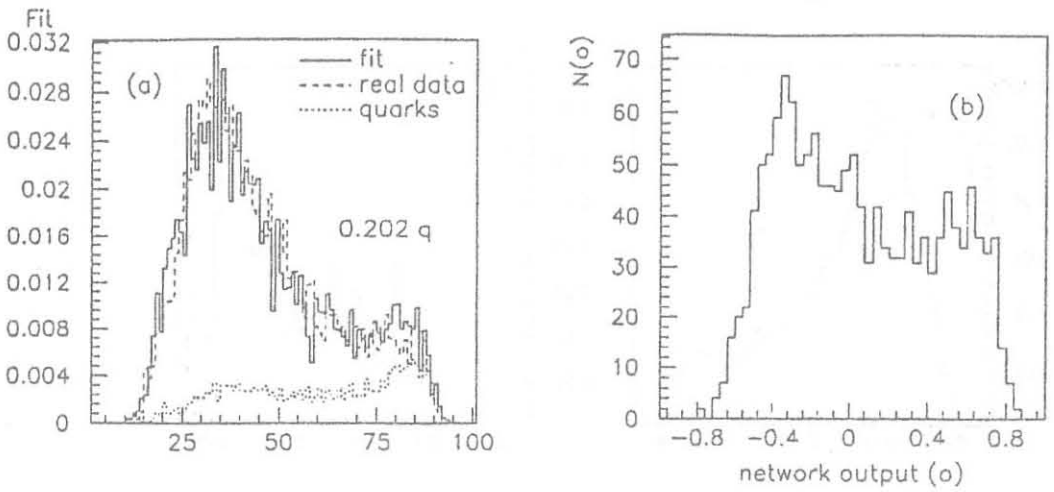

Figure 3: *(a) Output distribution from HERWIG data (80% quarks and 20% gluons) propagated through a PYTHIA-trained NN and fitted with PYTHIA distributions. The proportion of quarks extracted is displayed. (b) The distribution of B quarks propagated through the NN.*

the jet-20 data is composed primarily of gluons, with a small admixture of quarks. The expectation from QCD is 10-20% in this kinematical region (an attempt to extract the quark fraction is reserved for a future work). It is encouraging to note that the AMY dijets, which are expected to by exclusively quark jets, are interpreted primarily as such by the NN.

5. Classification with Fisher Discriminant

It is interesting to compare the previous results with a standard linear classifier like the Fisher Discriminant[11]. The discriminating power of this method was found to be worse than that of the NN applied to the same data. It is known that the Fisher Discriminant approaches the Bayes limit only if the input distributions are Gaussian and if the covariant matrices of the two samples are identical. One concludes that these conditions are not fulfilled in this application.

6. Conclusion

The effectiveness of the jet parameters proposed, in identifying parton ancestors on a jet-by-jet basis, has been investigated using a feed-forward NN. After optimization, the NN was seen to be equivalent to a Bayes classifier. A method for using the NN output distributions corresponding to simulated quark and gluon jets, to extract their relative proportions, was evaluated and model-dependence was estimated at 15%. Flavor-dependence was observed, but only for $E_t \leq 30GeV$. Real jet data from CDF and dijet data AMY were also propagated through the NN. A qualitative examination of the results was encouraging. Finally, a Fisher Discriminant was applied to the same data for comparison. Results were not as

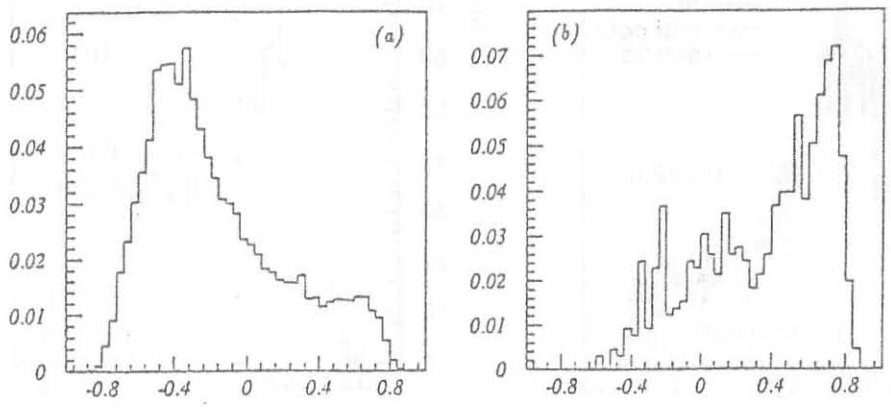

Figure 4: *Output distributions from (a) jet-20 data and (b) AMY data PYTHIA-trained NN*

good.

† *This work was funded in part by the Italian INFN and the US DOE.*

References

1. Z. Fodor, CERN-TH 6006/91; S. Kanda, S. Kim and K. Kondo, CDF/ANAL/JET/CDFR/959 July 19 (1990) L. M. Jones, *Phys. Rev.* **D42** (1990) 811; and refs. therein.
2. J. B. Gaffney and A. H. Mueller, *Nucl. Phys.* **B250** (1985) 109; G. Sh. Dzaparidze, *Z. Phys.* **C32** (1986) 59 and refs. therein
3. S. Bethke, CERN-PPE/91-36 (Feb. 26 1991); and refs. therein.
4. L. Lonblad, C. Petersen and T. Rongvaldsson, *Nucl. Phys.* **B349** (1991) 675; *Phys. Rev. Lett.* **65**; and refs. therein.
5. I. Csabai, F. Czako and Z. Fodor, CERN-TH.6038/91 (March 1991); T. AkeT. Akessonn and O. Barring, DELPHI note 90-59 PHYS 78 (12 Dec. 1990); L. Bellantoni et. al. CERN - PPE/91-80 (1991).
6. B. Denby et al. , *IEEE Trans. on Nucl. Sci.* **37** (1990) 248; P. Bhat et al., *Lund preprint LU TP 90 - 13* (1990).
7. H. Bengtsson and T. Sjostrand, *Comp. Phys. Comm.* **46** (1987) 43;
8. F. Fogelman Soulie *Artificial Neural Networks* Proc. of ICANN 1991, ed. T. Kohonen (Elsvier Sci. Publishers B.V.) (North Holland, 1991).
9. L. Garrido and V. Gaitan, U. Barcelona Preprint UAB-LFAE 91-04 (1991).
10. We are indebted to the AMY collab. for the use of their jet data.
11. R. A. Fisher, *Annals of Eugenics* **7** (1936) 179