UNIVERSITA' DI UDINE — ISTITUTO DI FISICA
Via Fagagna 208, I-33100 Udine (Italy)

Udine Report 92/08/AA

10 September 1992

Neural Networks for Physics Analysis in DELPHI

A. De Angelis and P. Del Giudice

Abstract

The use of Artificial Neural Networks in physics analysis by the DELPHI experiment at LEP is reviewed. DELPHI has used Neural Networks to tag the primary flavour of hadronic Z^0 decays, to enhance the sensitivity of the searches for new particles, to measure properties of quark and gluon jets, and to identify final state particles.

Presented at the Second Workshop on Artificial Neural Networks: From Biology to High Energy Physics, Marciana Marina, 18-26 June 1992. To be published in the Proceedings.

International Journal of Neural Systems, Vol. 3, (Supp. 1992) 000–000
Proceedings of the Neural Networks from Biology to High Energy Physics

NEURAL NETWORKS FOR PHYSICS ANALYSIS IN DELPHI

Alessandro De Angelis
Dipartimento di Fisica dell'Universitá di Udine and INFN
Via Fagagna 208, I-33100 Udine, Italy

and

Paolo Del Giudice
INFN Istituto Superiore di Sanitá
Viale Regina Elena 229, I-00161 Roma, Italy

Received (to be inserted
Revised by Publisher)

The use of Artificial Neural Networks in physics analysis by the DELPHI experiment at LEP is reviewed. DELPHI has used Neural Networks to tag the primary flavour of hadronic Z^0 decays, to enhance the sensitivity of the searches for new particles, to measure properties of quark and gluon jets, and to identify final state particles.

1. Introduction

The interest on Artificial Neural Networks (NN) is steadily increasing in High Energy Physics (HEP)[1]. Some 250 articles have been published in the last three years, mostly on *possible* applications of NN to our field.

The term "neural" is somehow misleading for the use that physicists do of NN. In general, Neural Networks in HEP do not pretend to be a paradigm of the way of operating of the brain, but simply a function characterized by a large number of inputs and a high degree of interconnection between them. In other words, we use NN to draw effectively nonlinear hypersurfaces to separate events according to their different features, and thus allow their classification.

Ten years ago, most HEP experiments were using bubble chambers as detectors. Physicists were analyzing events having in front many different inputs, and no general recipe for classifying an event. Many attempts of writing software for automatic classification of events failed, due to the high variety of information (qualitative and quantitative) to be processed. Neural Networks were a break-through for the task of automatic event classification, allowing in many cases a robust separation. Nonlinearity allows the access to extra degrees of freedom with respect to the standard linear multidimensional analyses.

Among the different architectures of NN, Feed-Forward had a great success due to its simplicity and its adaptivity. The DELPHI experiment at the LEP e^+e^- collider started since intervals of rapidity the distributions of multiplicity are usually analyzed in terms of normalized factorial moments. Given an experimental distribution of particles in the rapidity interval from $-Y/2$ to $Y/2$, the interval Y is divided into M equal subintervals, each of size $\delta y = Y/M$. If N is the number of particles in the whole rapidity interval and n_m the number of particles in the m-th bin ($m = 1...M$), the factorial moment of (integer) rank j of the distribution is defined by

$$F_j(\delta y) = \frac{M^{j-1}}{< N >^j} < \sum_{m=1}^{M} n_m(n_m - 1)...(n_m - j + 1) > \tag{4}$$

where the averages are taken over many events. The factorial moment of rank j for a rapidity interval δy selects events with j particles or more in at least one bin and is sensitive to events with density fluctuations in rapidity[12].

It was suggested[13] that $b\bar{b}$ events have higher values of factorial moments than events coming from the hadronization of a lighter quark. The difference we observe on the second factorial moment is lower than what predicted by JETSET PS (Fig. 2).

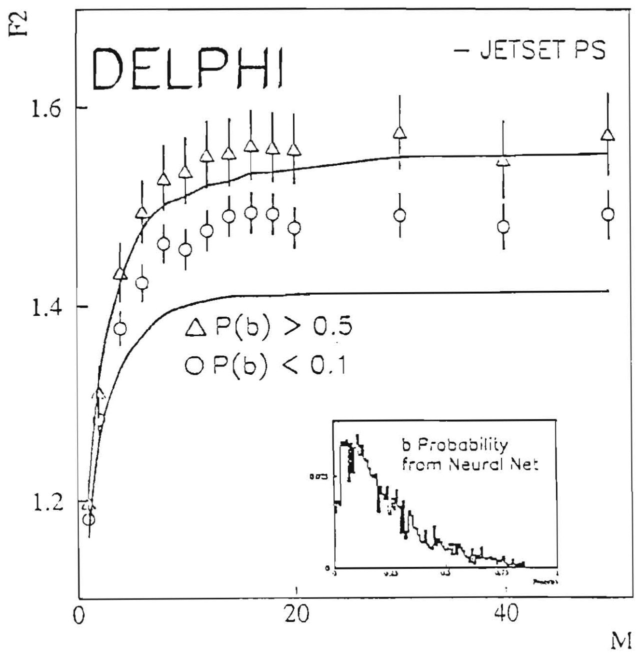

Fig. 2. Second factorial moment of the rapidity distribution, in the rapidity range between -2 and +2, for events with probability of coming from a $b\bar{b}$ pair larger than 0.5 (triangles) and smaller than 0.1 (circles) versus the number M of divisions of the rapidity interval. Comparison with JETSET PS (solid lines).

2.2.3. *Bose-Einstein Interference*

An enhancement in the production of pairs of pions of the same charge and similar momenta produced in high energy collisions is attributed to Bose-Einstein statistics appropriate to identical pion pairs.

The enhanced probability for emission of two identical bosons is studied in general through the two-particle correlation function, R, as a function of the 4-momentum difference Q. The correlation function is often parametrized by the function:

$$R(Q) = 1 + \lambda e^{-r^2Q^2} \tag{5}$$

where the parameter r gives the source size and λ measures the strength of the correlation between the pions, being 0 for a completely coherent source and 1 for a completely incoherent one.

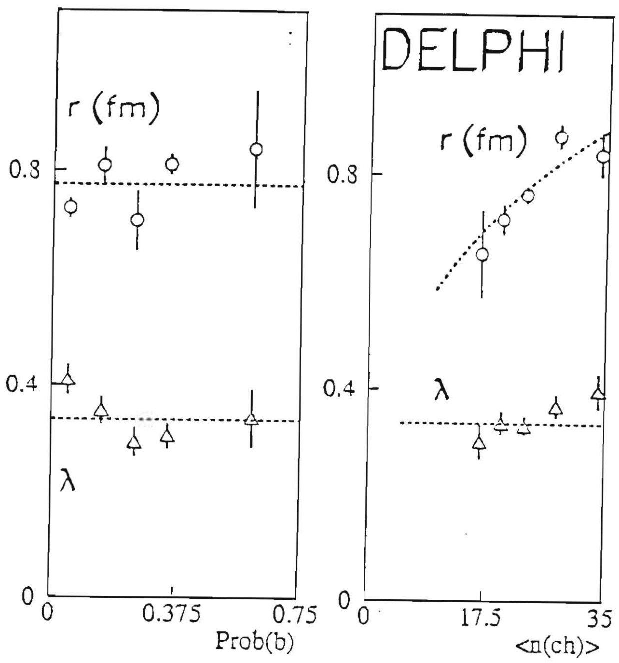

Fig. 3. Left: Radius r and coherence parameter λ, versus the probability of belonging to the class of $b\bar{b}$ events. Right: r and λ, for different multiplicities.

An expression slightly different from (5) is often used in the literature to analyze the experimental data:

$$R(Q) \quad = \quad N(1 + \delta \cdot Q)(1 + \lambda e^{-r^2Q^2}) \tag{6}$$

where the factor $(1 + \delta \cdot Q)$ is introduced to take into account possible long-range momentum correlations in the form of a slow rise, and N is a normalization factor.

In the present study, pairs of oppositely charged particles have been taken as a reference sample.

The bias due to residual correlations in the reference sample has been removed by dividing the ratio obtained from the data by the equivalent ratio from the simulation. This normalized ratio for the unlike charged reference sample has been fitted to form (6) for each bin of b purity. The results of the fit to r and λ are displayed in Fig. 3 (left), where the averages (consistent with[14]) are marked as dashed lines.

the beginning of its data-taking to explore the potentialities of such networks in the event classification task.

In particular, DELPHI explored four topics:

- The classification of the hadronic decays of the Z^0 according to their native flavour.
- The separation of jets coming from the hadronization of a quark from those coming from the hadronization of a gluon.
- The searches for new particles.
- The identification of secondary particles in the final states.

In this paper, we will shortly review the activity of DELPHI on these topics.

2. Neural Networks for Flavour Classification

The difficulties in the classification of hadronic events according to their parent quark flavour can be overcome by utilizing multi-dimensional variables for the separation[2]. Among the multi-dimensional classifiers, Feed-Forward Neural Networks appear to be a good candidate for complicated problems such as this. We use them to map a set of variables calculated from the event onto a feature space in which the different species are well separated.

The possibility to use a NN for hadronic event classification was explored in Ref. [3], in which the problem of separating Z^0 decays into $b\bar{b}$ pairs was considered. The result of this study was that, in the case of a perfect detector, a separation could be achieved with a higher efficiency than with respect to traditional separation variables[2]. Further studies[4] demonstrated that, also in the presence of detector effects, NNs could be a useful tool for the classification of $b\bar{b}$ events, and preliminary results on real data have been recently presented[5,6].

The physics motivation of a careful analysis of the hadronic branching fractions of the Z^0 was in the beginning to measure very accurately the branching ratio into $b\bar{b}$ pairs. The decay of the Z^0 into $b\bar{b}$ can in fact happen via a loop involving the production of a virtual $t\bar{t}$ pair. Thus, the relative amplitude $\Gamma_{b\bar{b}}/\Gamma_h$ (where Γ_h is the total hadronic width of the Z^0) decreases as the mass of the top quark increases. If one could identify *perfectly* 300,000 hadronic decays of the Z^0, the mass of the top quark would be known at ± 25 GeV.

DELPHI has recently shown[7] that a NN can be used to classify effectively decays of the Z^0 into $b\bar{b}$ and $c\bar{c}$ pairs. The robustness of the separation against a wide range of systematics related to model-dependence of the classification has been investigated. As a result, it has been possible to measure, from the data collected by the DELPHI detector at LEP during 1991, the rates of the hadronic decays into $b\bar{b}$ and $c\bar{c}$.

2.1. *Hadronic Branching Fractions of the Z^0*

The first measurement by DELPHI[5] was performed by coupling four single-output feed-forward neural networks, specialized 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[8].

Eighteen variables, mainly related to the shape of the event, were used as an input for the separation. 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[5] for details on the training of the networks, based on back-propagation, and on their architecture.

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

$$\Re^{(i)}(t) = \sum_j \beta_j^{(i)} a_j^{(i)}(t) , \tag{1}$$

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 a set of simulated events independent from the sample used for the training (*test sample*). All distributions were normalized 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 $\chi^2 = \sum_i < \vec{\beta}^{(i)} - \vec{\beta}^* | C^{(i)-1} | \vec{\beta}^{(i)} - \vec{\beta}^* >$, where $C^{(i)}$ is the covariance matrix in the fit from the i-th network, was minimized with respect to the unknown branching fractions $\vec{\beta}^*$, under the constraint that the sum of the branching fractions is equal to 1. This led to the central determinations and to the statistical errors.

The study of systematics kept into account the uncertainties in the best tuning of parameters in JETSET PS as parametrized in[9]. In addition, a detailed study of the effect of fragmentation parameters was done, to obtain the systematic errors.

The final results were

$$\begin{aligned}
\Gamma_{u\bar{u}+d\bar{d}}/\Gamma_h &= 0.417 \pm 0.015(stat) \pm 0.058(sys) \\
\Gamma_{s\bar{s}}/\Gamma_h &= 0.233 \pm 0.016(stat) \pm 0.051(sys) \\
\Gamma_{c\bar{c}}/\Gamma_h &= 0.139 \pm 0.010(stat) \pm 0.058(sys) \\
\Gamma_{b\bar{b}}/\Gamma_h &= 0.211 \pm 0.006(stat) \pm 0.020(sys) .
\end{aligned}$$

Four problems were left to be solved:

(i) Correlations between the four networks. The branching fractions were extracted by constructing an overall χ^2 from the four networks, keeping into account the covariances only at the first order.

(ii) Correct estimate of systematics, affecting the determination of the a_j's in (1). 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*.

The radius r of the source of the hadronization and the coherence parameter λ do not display a clear dependence on the percentage of b events present in the sample, although the results are compatible with a slight dilution of the effect for b events.

The average number of particles produced via strong interaction in $b\bar{b}$ events should be smaller than for light quarks. One could thus expect a smaller radius of the hadronization region. This is not observed, although, when one studies the dependence of the radius on the multiplicity, a trend consistent with being proportional to $< n >^{1/3}$ (Fig. 3, right) is seen.

2.3. *Conclusions and Possible Improvements*

In conclusion, the flavour classification problem received effective solutions from Neural Networks, although the systematic error from the method seems of the order of 5% in the case of the estimate of the branching fraction into $b\bar{b}$ pairs, in the order of 15% for $c\bar{c}$ pairs. The precision achieved is thus not such that one can constrain the mass of the top quark.

Other analyses from the same collaboration[15,16], based on simulated data, do not improve the efficiency for the separation.

Possible improvements can be obtained form a better choice of the input variables, and from modifications of the training. A detailed analysis of these two topics was done in[17].

2.3.1. *A new kind of input variables*

A common difficulty that one faces in choosing input variables for the network is how to use single particle information. In fact the number of input variables for the network is usually held fixed and retaining only the information from the most energetic particles could be neither sufficient nor appropriate.

An alternative strategy for including single particle information has been explored, based on a probabilistic characterization of the events. Each event undergoes a hypothesis test for the distribution in the event of a single particle variable against reference distributions of the same variable for the different classes to be discriminated, yielding probabilities for the considered event to belong the each one of the classes.

There are physical reasons to believe that the impact parameter contains relevant information to tag the b quark but the available measurements are such that its distributions for the b and non b classes are largely overlapped. So we decided to apply the hypothesis test strategy to the impact parameter, trying in this way to exploit also shape differences between each of the reference distributions typical of each class and the single event ones. We selected a subset of the Monte Carlo sample (1154 events) to build the reference distributions of the impact parameter for the b class, c class and light quarks class. Then for each event the impact parameter distribution is compared with the reference ones and the three probabilities p_b, p_c and p_l are calculated by using the Kolmogorov-Smirnov test[18]. This variable demonstrates to be the most effective for classification of b quarks.

2.3.2. *Modified Training Algorithms*

Possible ways to modify the algorithm have been explored, to influence the way the network learns from the examples, trying to obtain somewhat controlled output patterns and performances. The guiding concept of all the considered modifications is the definition of a "quality" criterion for the events in the training data set and the corresponding definition of the "exceptional" cases. We consider two different ways to define this criterion:

• the training data set is sorted in descending order according to the values of $\frac{p_b}{p_c}$ for the b events and $\frac{p_c}{p_b}$ for the non b ones. The lower the position of the event in the sorted file, the better its "quality";

• the "goodness" of each event is defined dynamically by the network as the events whose outputs satisfy a preselected requirement defined by the problem at hand.

Whatever criterion one chooses, it allows to define alternative algorithms, amounting essentially to a redefinition of the error function. The first criterion is applied to two cases:

• the input data set is enlarged during the training, including more and more "exceptional" events ("pedagogical training")

• the input data set is modified, changing to 0 the desired output of the last 500 "exceptional" b events in the training set ("simplified training"). Such a drastic change is suggested by the observation that "exceptional" events of each class, which are responsible for the difficulty of the problem, resemble the "good" events of the other class so much that, letting the "exceptional" b to be badly learned results in a simplification of the problem, possibly leading to a better understanding of the unchanged events.

The second criterion makes the network itself decide which events are privileged. One can then influence the learning by defining a modified error function

$$E = \frac{1}{N_{examples}} \left(\alpha \sum_{privileged} (O - \overline{O})^2 + \sum_{others} (O - \overline{O})^2 \right) .$$

Acting on the coefficient α, it is possible to force learning some patterns better than others. This modification has been applied to a network already trained with the simplified learning algorithm, defining as privileged events those yielding an output greater than 0.9 and choosing $\alpha = 5$.

No unambiguous improvements come from this kind of modified algorithms, although a wise use of them can allow to increase efficiency *or* purity in a controlled way.

3. Classification of Jets coming from Quarks and Gluons

Many interesting experimental tests of QCD require at least a fair knowledge of the parton origin of a hadronic jet. In some cases, e.g. measurement of the string effect in 3-jet events or the gluon-gluon coupling in multijet events, it is sufficient to distinguish gluon from quark jets.

Normally, quark jets are tagged by means of the presence of a lepton radiated from a semileptonic decay (in the case of heavy flavours), or by the fact that jets originated from quarks have in average a larger energy.

Neural Networks allow a classification insensitive to the jet energy, i.e., exploiting only the fragmentation differences between quarks and gluons [19].

In detector simulated Monte Carlo data, the identification of the two jets originated from the primary partons in 4-jet events is about 10% better than with the jet energy classification. By making use of this approach, DELPHI has improved significantly the evidence for triple-gluon vertices in hadronic decays of the Z^0.

4. Searches for New Particles

NN can enhance the sensitivity in the searches for new particles.

A search for the Minimal Standard Model Higgs boson, through the reaction $e^+e^-\rightarrow H^0\nu\bar{\nu}$, using the data collected in 1990 by the DELPHI detector at LEP, has been made by means of Neural Networks[20].

The basic idea is to teach a NN to separate the Higgs signal from the background, using simulated data, and then to use the NN output variables for the definition of the selection criteria. "Standard" selection variables have been coupled to the NN output for the task of the separation.

The analysis based on NN is much better than the standard one[21] for masses around 50 GeV, being the efficiency for Higgs detection almost doubled.

The technique used allows to reach good detection efficiencies and better limits than those reached by means of standard analysis techniques, thanks to the advantages given by the possibility to use information which would be, otherwise, very difficult to use. The price to pay is the fact that the classification method is not completely transparent.

5. Identification of Final-State Particles

Finally, DELPHI has used NN for improving the efficiency of the capability of the detector to identify final state particles.

Often the particle identification is the result of the combination of different inputs from different subdetectors. The use of NN is under study, to improve the efficiency in the identification of electrons[22], muons[23] and hadrons from the Barrel Ring Imaging Cherenkov[24].

6. Conclusions

After three years of exploration of the potentialities of Neural Networks for physics analysis, the DELPHI experiment is publishing the first results. Significant results were obtained, expecially in the task of automatic event classification, and significant developments are foreseen.

References

1. C. Bortolotto et al., *Int. J. Mod. Phys.* **C3** (1992) 733.
2. R. Marshall, *Zeit. Phys.* **C26** (1984) 291.
3. L. Lönnblad et al., *Nucl. Phys.* **B349** (1991) 675.
4. C. Bortolotto, A. De Angelis and L. Lanceri, *Nucl. Instr. Meth.* **A306** (1991) 459.
   L. Bellantoni et al., *Nucl. Instr. Meth.* **A310** (1991) 618.
5. C. Bortolotto et al., Proceedings of the Workshop on "Neural Networks, From Biology to High Energy Physics", Isola d'Elba (Italy), June 1991, p. 445.
6. P. Henrard (ALEPH), presented at the 4th Symposium on Heavy Flavour Physics, Orsay, June 1991.
   B. Brandl (ALEPH), Heidelberg preprint HD-IHEP 92-01, to be published in the Proceedings of the Second International Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, L'Agelonde, January 1992.
7. P. Abreu et al. (DELPHI), "Classification of the Hadronic Decays of the Z^0 into b and c Quark Pairs using a Neural Network", CERN-PPE/92-151 (September 1992), submitted to Phys. Lett. B.
   P. Eerola et al. (DELPHI), these proceedings.

8. A. De Angelis, "Hadronic Branching Fractions of the Z^0 boson", Proceedings of the International Symposium on Multiparticle Dynamics, Wuhan (China), September 1991, p. 57.
9. M.Z. Akrawy et al. (OPAL), *Zeit. Phys.* **C47** (1990) 505.
10. W. de Boer et al., IEKP-KA/91-07, Karlsruhe, June 1991.
11. P. Abreu et al. (DELPHI), *Zeit. Phys.* **C50** (1991) 185.
12. P. Abreu et al. (DELPHI), *Phys. Lett.* **B247** (1990) 137.
    A. De Angelis and N. Demaria, Proc. of the Ringberg Workshop on "Multiparticle Production Fluctuations and Fractal Structure", Rindberg, Germany, June 1991, eds. R. C. Hwa, W. Ochs, N. Schmitz, World Scientific, Singapore, p. 1.
13. A. Giovannini, R. Ugoccioni and L. van Hove, Proceedings of the Int. Workshop on Correlations and Multiparticle Production (CAMP), Marburg 1990.
14. P. Abreu et al. (DELPHI), *Phys. Lett.* **B286** (1992) 201.
15. M. Los and N. DeGroot, Proceedings of the Workshop on "Neural Networks, From Biology to High Energy Physics", Isola d'Elba (Italy), June 1991, p. 459.
16. O. Botner and K. Kulka, DELPHI internal note, CERN 1992.
17. P. Branchini et al., INFN-ISS 92/1 (January 1992), to be published in the Proceedings of the Second International Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, L'Agelonde, January 1992.
18. see e.g. Eadie et al., "Statistical Methods in Experimental Physics", North Holland 1971.
19. O. Barring et al. (DELPHI), Contributed paper to the XXVI International Conference on High Energy Physics, Dallas 1992.
20. A. Petrolini, *Int. J. Mod. Phys.* **C3** (1992) 611.
21. P. Abreu et al. (DELPHI), *Nucl. Phys.* **B373** (1992) 3.
22. F.L. Navarria, DELPHI internal note, CERN 1991.
23. M. Los and N. DeGroot, these proceedings.
24. N. DeGroot, M. Los and S. Tzamarias, Proceedings of the Workshop on "Neural Networks, From Biology to High Energy Physics", Isola d'Elba (Italy), June 1991, p. 561.