DEUTSCHES ELEKTRONEN-SYNCHROTRON

DESY 93-167
November 1993

Application of Neural Networks in the ARGUS-Experiment for the Analysis of B- and Tau-Physics

M. Joswig, H. Kolanoski, S. Kublun, H. Thurn, D. Wegener, S. Westerhoff

Institut für Physik, Universität Dortmund

ISSN 0418-9833

NOTKESTRASSE 85 - 22603 HAMBURG

To be sure that your preprints are promptly included in the HIGH ENERGY PHYSICS INDEX, send them to (if possible by air mail):

DESY **Bibliothek** **Notkestraße 85** **22603 Hamburg** **Germany**	**DESY-IfH** **Bibliothek** **Platanenallee 6** **15738 Zeuthen** **Germany**

DESY 93-167 ISSN 0418-9833
November 1993

APPLICATION OF NEURAL NETWORKS IN THE ARGUS EXPERIMENT FOR THE ANALYSIS OF B- AND TAU-PHYSICS*

M. JOSWIG, H. KOLANOSKI, S. KUBLUN, H. THURN, D. WEGENER and S. WESTERHOFF
Institut für Physik, Universität Dortmund,
D-44221 Dortmund, Germany

Presented by H. KOLANOSKI at the Third International Workshop
on Software Engineering, Artificial Intelligence and Expert Systems
for High Energy and Nuclear Physics
October, 4 - 8 1993, Oberammergau, Oberbayern, Germany

ABSTRACT

We present analyses of e^+e^- annihilation data taken with the ARGUS detector in the energy range of the Υ resonances. The selection of $\Upsilon(4s)$ decays and leptonic τ decays was studied with feed-forward multilayer networks using the backpropagation algorithm for the training. The correlated decay of $\tau^+\tau^-$ pairs was analysed with a feed-forward network to determine a weak coupling constant.

1. Introduction

Until the beginning of this year the ARGUS experiment[1] took data at the e^+e^- storage ring DORIS II in the energy range of the Υ resonances ($E_{e^+e^-} \approx 10$ GeV). The most important physics contributions were made to the understanding of the 3rd fermion generation studying b quark and τ lepton decays.

In this paper we present the application of neural network algorithms for the selection of B decays[2] and τ events[3] and for the determination of an electroweak parameter in τ decays[4]. The studies reported here used feed-forward nets with at most 2 hidden layers and the backpropagation algorithm as implemented in the JETNET program[5]. The training was done mainly with simulated events and rather standard settings for the learning parameters (step size, inertia term, updating frequency).

2. Selection of $\Upsilon(4s)$ Events

The properties of the b-quarks have been extensively studied by analysing the decays of B mesons[6]. At e^+e^- colliders associated $B\overline{B}$ production is observed in decays of the $\Upsilon(4s)$ resonance:

$$e^+e^- \rightarrow \Upsilon(4s) \rightarrow B\overline{B}.$$

In competition to the resonance production is the production of a continuum of light $q\overline{q}$ pairs. At the $\Upsilon(4s)$ the signal of $B\overline{B}$ events to the background of continuum

*This work was supported by the German Bundesministerium für Forschung und Technologie under contract number 056DO57P.

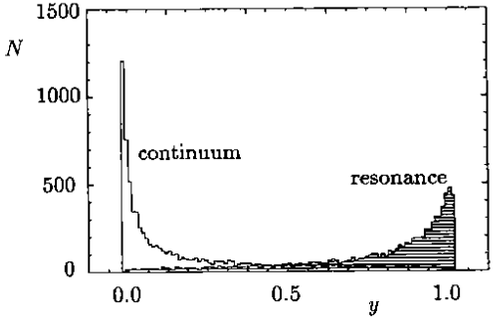

Figure 1: The two output distributions for simulated continuum and resonance events.

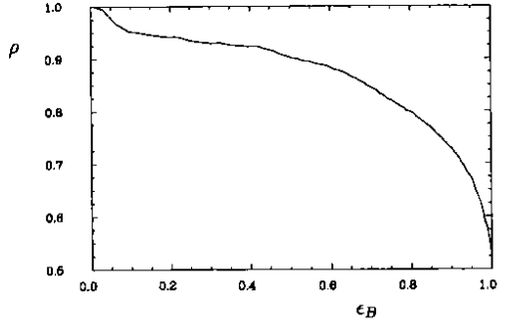

Figure 2: Purity versus selection efficiency for $B\overline{B}$ events. The 'purity' is defined as the fraction of real $B\overline{B}$ events in the sample.

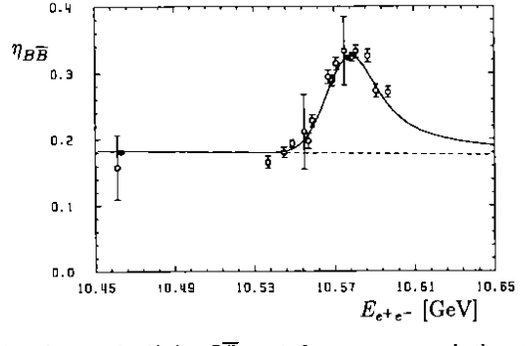

Figure 3: The fraction of events classified as $B\overline{B}$ events for an energy scan in the region of the $\Upsilon(4s)$ resonance. The full curve is a fit to the resonance plus the continuum (dashed line: continuum only).

events is about 1:4. We have studied the separation of the resonance decays from the continuum using a feed-forward neural network together with backpropagation training.

The most characteristic difference between both event classes is found in the event topology: Since the $\Upsilon(4s)$ resonance is just above the $B\overline{B}$ threshold the B mesons are produced nearly at rest and therefore decay isotropically. The continuum events, on the other hand, have dominantly back-to-back two-jet topologies. Therefore topological quantities, like sphericity or Fox-Wolfram moments, are important input variables for the net. Other used variables exploit the somewhat different multiplicities (higher in resonance decays), the occurrence of leptons and kaons in the weak decays of B mesons and kinematical differences (e.g. the maximal momentum of a particle in resonance decays is restricted to half the B mass). Altogether 20 characteristic variables were chosen as network inputs, normalized such that each variable approximately had values in an unit interval.

The evaluation of the results for various network configurations led finally to a 20-20-1 configuration. This net was trained with simulated events to the target output values $y = 1$ for the resonance reaction and $y = 0$ for the continuum. After about 300 learning epoches with about 30000 events typical output distributions as in Fig.1 were obtained. Varying the y cut the efficiency for $B\overline{B}$ selection could be tuned as a function of the remaining continuum background (Fig.2). A cut at $y = 0.5$ yielded a selection efficiency of $(84.6 \pm 0.5)\%$ and a background rejection of $(80.0 \pm 0.6)\%$. This has been applied to experimental data from a scan over the resonance. A signal-to-background ratio of about 1:1 at the resonance peak could be reached as shown in Fig.3.

The peak cross section obtained with the neural network is $\sigma = (0.87 \pm 0.01)\ nb$. With the standard procedure of fitting the resonance on top of the full (unsuppressed) background $\sigma = (0.86 \pm 0.03)\ nb$ was obtained which has a substantially larger error than with the network selection. However, the background suppression introduces a larger dependence on the simulation leading to an additional systematic error of about 3%.

3. Selection of Leptonic τ Decays

This analysis aimed at the determination of the product of the leptonic branching ratios of τ leptons, $B_e \cdot B_\mu$, studying the τ production and decay process

$$e^+e^- \to \tau^+\tau^- \to \mu^+\nu_\mu\overline{\nu}_\tau\, e^-\overline{\nu}_e\nu_\tau.$$

The network had to recognize

- 2 leptons in the event,
- missing energy and momentum due to the four undetected neutrinos

and to distinguish the wanted reaction from background with similar signatures. The most difficult background came from other τ decays and two-photon reactions.

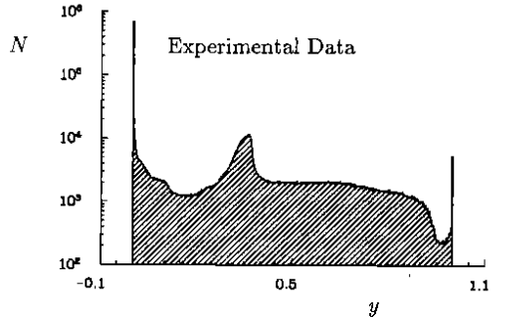

Figure 4: Network output for experimental data (leptonic τ decay candidates).

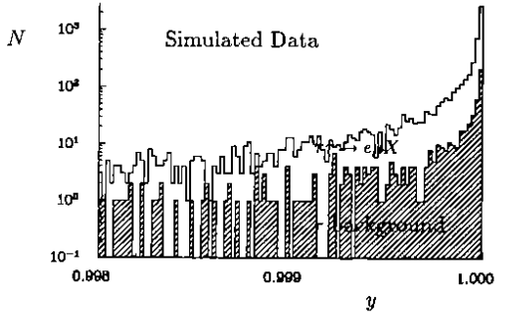

Figure 5: Network output near $y = 1$ for simulated leptonic τ decays (open histogram) and simulated background from other τ decays (shaded histogram).

The final analysis was done using a 15-15-1 feed-forward net. The 15 input variables were the momentum components of the two tracks, associated particle identification quantities derived from shower energies, muon chamber hits, specific energy loss and time-of-flight measurements, and in addition kinematical quantities for the whole event, like missing mass and the direction of the event momentum. The necessary number and the optimal choice of input variables was determined from studies of different network configurations. These studies also showed a major advantage if the variables of each lepton species were assigned to specific input nodes.

The net was trained using 5000 simulated events composed of the studied reaction and the background in approximately the expected composition. After about 500 learning epoches a selection efficiency above 90% for the leptonic τ de-

cays was reached. However, the network output distribution for the experimental data in Fig 4 shows no clear separation between signal and background. The bumpy structure of the distribution was found to be due to the different background contributions, e.g. the bump near $y = 0.4$ is due to the two-photon reaction.

It turns out that signal and background cannot be completely separated, which was also found in conventional analyses of the same reaction[7]. Figure 5 shows that even near $y = 1$ the background from other τ decays persists. Varying the cut on the output variable y a reasonable signal-to-background ratio was found for $y_{cut} = 0.985$. With this cut the selection efficiency was

$$\epsilon_{sel} = 0.247 \pm 0.003$$

for a remaining background of 9%. The final result was

$$B_e \cdot B_\mu = 0.0295 \pm 0.0007 \pm 0.0015.$$

The two errors are statistical and systematical, respectively. The systematic error is estimated from the variation of $B_e \cdot B_\mu$ when changing the y-cut and using different network input variables.

In a conventional analysis[7] using one- and two-dimensional selection cuts an efficiency $\epsilon_{sel} = 0.147 \pm 0.003$ was reached for a remaining background of only 2%. Since the background is in this case much smaller the two efficiency numbers are not quite comparable. However, it seems as if both methods could yield similar signal-to-background ratios. Nevertheless, one may see advantages in using a neural network:

- The cuts in the conventional analysis have been fine-tuned in a tedious and time-consuming way. Network training offers a more automatic optimisation of cuts (and takes into account multidimensional correlations).
- The evaluation of inconsistencies between Monte Carlo simulations and the data can be made more systematically by varying the y-cut.

The systematic uncertainties due to incorrect simulations affect both the conventional and the network approach. It should be emphasized that a net training with an incorrectly simulated data set does not necessarily lead to a wrong data analysis as long as the efficiency corrections are done for this specific (may be not optimal) network correctly.

4. Determination of Electro-Weak Coupling Parameters from $\tau^+\tau^-$ Decay Correlations

In this work we investigated possibilities of extracting parameters of statistical distributions by the neural network approach. In contrast to the above discussed examples this is not a classification (yes-no) problem but the aim is to calculate a continuous number. In case of a linear dependence of the distribution $p(\vec{x})$ on the

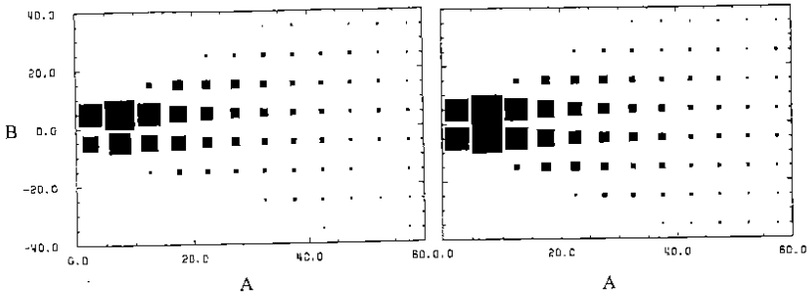

Figure 6: Distribution of events as a function of the A and B terms simulated for γ_{AV}^2=1 (left) and γ_{AV}^2=0 (right).

parameters λ_i,

$$p(\vec{x}|\vec{\lambda}) = \sum_i \lambda_i p_i(\vec{x}),$$

the distribution of the net output can be linearly composed of the output distributions for the different $p_i(\vec{x})$. Thus the λ_i can be determined from a fit of simulated output distributions to the output distribution of the data. Using neural networks compared to, e.g., a likelihood analysis has the advantage that the $p_i(\vec{x})$ have not to be explicitely known but the net can be trained with simulated data.

This method has been applied to the determination of the vector-axialvector interference in correlated hadronic decays of τ pairs. The angular distribution of the final state in

$$e^+e^- \to \tau^+\tau^- \to \pi^+\pi^0\overline{\nu}_\tau\pi^-\pi^0\nu_\tau$$

can be described by[8]

$$p(\vec{\alpha}|\gamma_{AV}^2) = A(\vec{\alpha}) + \gamma_{AV}^2 B(\vec{\alpha})$$

where $\vec{\alpha}$ is a set of 11 kinematical variables and γ_{AV} is the normalized product of the vector and axialvector charged current couplings. The B term can be positive and negative which generates according to the equation above a higher event density for positive B and a lower for negative B (Fig.6).

The distribution for any γ_{AV}^2 can be obtained as a linear combination of the two distributions with $\gamma_{AV}^2 = 0$ and 1:

$$p(\vec{\alpha}|\gamma_{AV}^2) = \gamma_{AV}^2\, p(\vec{\alpha}|1) + (1-\gamma_{AV}^2)\, p(\vec{\alpha}|0) = \gamma_{AV}^2(A+B) + (1-\gamma_{AV}^2)A$$

The idea is now to train a feed-forward net to distinguish the distributions $p(\vec{\alpha}|1)$ and $p(\vec{\alpha}|0)$. This can be achieved by requiring the target output (1 output node) to be $y = 1$ and $y = 0$ for $p(\vec{\alpha}|1)$ and $p(\vec{\alpha}|0)$, respectively. After such a net training output distributions for $p(\vec{\alpha}|1)$ and $p(\vec{\alpha}|0)$ are obtained as shown in Fig.7a. The two distributions are not very distinguished due to the relatively small asymmetry observed in Fig.6.

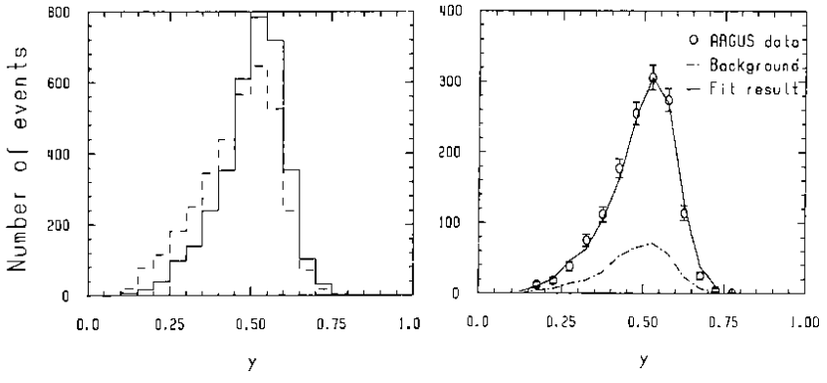

Figure 7: a) Network output distribution for simulated events with γ^2_{AV}=1 (full line) and γ^2_{AV}=0 (dashed line).
b) Network output distribution for the 1412 ARGUS events (points with error bars) and the background. The curve is the result of the fit described in the text.

There are two additional experimental complications to be taken into account: Because of the escaping neutrinos the τ direction can only be reconstructed up to a two-fold ambiguity (and even that is only true if one neglects radiative corrections). This yields for each event two sets of kinematical variables and hence also two (A,B) sets. We have finally used a 4-10-4-1 net with A, B, A', B' as the four inputs (Fig.7a was obtained with this net). The second complication is a quite substantial background of about 25%, mainly from other τ channels. This background could be simulated and was also fed through the net to generate its output distribution.

Figure 7b shows the network output distribution for the selected 1412 events together with the simulated, properly normalized background. The data distribution was now fit by the two reference distributions in Fig.7a and the background distribution. The latter was held fixed in the fit. The only free parameter, $|\gamma_{AV}|$, was determined to be:

$$|\gamma_{AV}| = 0.96 \pm 0.05 \pm 0.04 \quad \text{(prelim.)}$$

The errors are statistical and systematical, respectively. The systematical error is mainly due to the uncertainty in the background determination. No systematic effects specific to the use of neural networks have been identified.

The network result can be compared to a likelihood analysis using the complete (but uncorrected) matrix element. The preliminary result from this analysis is $|\gamma_{AV}| = 1.00 \pm 0.03 \pm 0.03$. With the likelihood method the detector and radiative effects cannot easily be included and can only be corrected for by detailed simulation studies. In this respect the network approach is advantageous since all simulated

effects are naturally included by the training. It is under investigation if the statistical error in the network result can be decreased further, e.g. by improving the statistics of the distributions used in the fit and/or by using the kinematical variables explicitely rather than the A and B terms.

5. Summary and Conclusion

We have successfully applied feed-forward networks to select $\Upsilon(4s)$ events and leptonic τ decays. The advantage of using neural networks is most obvious for the selection of $\Upsilon(4s)$ events. These events differ in each of the chosen variables (multiplicities, topology, ...) often only by a small amount and therefore a symbiotic view of all the variables is particularly important. On the other side, in the selection of leptonic τ decays the particle identification plays the dominating role and the decision seems to be to less an extend multi-dimensional. In this case one may consider to put more basic information on particle identification onto the net rather than preprocessed classifiers.

The third example shows that neural networks can be used to determine linear coefficients of distributions. In this method the necessity of fitting the output distributions has little 'neural appeal'. More studies are necessary to search for networks which could provide more directly the solutions and maybe also the generalisation to non-linear parameters.

In the studied applications the neural networks have provided satisfactory problem solutions, sometimes better than what could be reached by other methods. The identified advantages of the neural analysis approach are:

- quasi-automatic learning in a multi-dimensional space;
- natural use of simulated or measured data: the distributions have not to be explicitely known;
- the signal-to-background ratio can be smoothly adjusted by varying the cut on the network output;
- this cut on the output allows also to check systematically differences between simulation and experimental data;
- by the intrinsic parallelism of neural algorithms neural networks are optimal for use on the trigger level.

A final remark: from our experience neural network classification is not less transparent than any other multi-dimensional analysis technique. There are methods to analyse the network behaviour and to visualize what a network is doing. One can for example find out what the relevance of the different input variables is. In most cases one finds no surprises: a good physicist would choose the same variables.

References

1. ARGUS Coll., H. Albrecht et al., Nucl. Inst. Meth. A275 (1989) 1.
2. M. Joswig, diploma thesis, University of Dortmund, 1993.
3. S. Westerhoff, diploma thesis, University of Dortmund, 1993.
4. S. Kublun, diploma thesis, University of Dortmund, 1993.
5. L. Lönnblad, C. Peterson and T. Rögnvaldsson, *'Pattern recognition in high-energy physics with artificial neural networks: JETNET 2.0'*, Comput. Phys. Commun. 70 (1992) 167.
6. H. Schröder, *'Physics of the B Mesons'*, Rep. Prog. Phys. 52 (1989) 765.
7. ARGUS Coll., H. Albrecht et al., Z. Phys. C53 (1992) 367.
8. H. Thurn and H. Kolanoski, submitted to Z. Phys. C, (DESY 93-056), 1993.