DEUTSCHES ELEKTRONEN-SYNCHROTRON DESY

DESY 90-144
LU TP 90-13
November 1990

Using Neural Networks to Identify Jets in Hadron-Hadron Collisions

P. Bhat

Fermi National Accelerator Laboratory, Batavia, USA

L. Lönnblad

Department of Theoretical Physics, University of Lund

K. Meier

Deutsches Elektronen-Synchrotron DESY, Hamburg

K. Sugano

Argonne National Laboratory, Argonne, USA

ISSN 0418-9833

NOTKESTRASSE 85 · 2 HAMBURG 52

DESY 90-144
LU TP 90-13
November 1990

ISSN 0418-9833

Using Neural Networks to Identify Jets in Hadron-Hadron Collisions

Pushpalatha Bhat
Fermilab, P.O. Box 500, Batavia, IL 60510,USA

Leif Lönnblad
Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Karlheinz Meier
DESY, Notkestraße 85, D-2000 Hamburg 52, Germany

Katsuhito Sugano
Argonne National Laboratory, Argonne, IL 60439, USA

Contribution to the proceedings of the *1990 Summer Study on High Energy Physics - Research Directions for the Decade, Snowmass, Colorado, June 25 - July 13, 1990.*

Abstract:

We investigate the use of neural network techniques to discriminate between quark and gluon jets in Monte Carlo generated $p\bar{p}$ collisions. We also investigate the use of this technique to distinguish hadronically decaying W and Z bosons from QCD background. In the latter case we find that the signal-to-background ratio may be increased by more than a factor of 10 using a neural network trigger.

1 Introduction

Copious production of hadronic jets is the dominant feature of high transverse momentum processes in $p\bar{p}$ collisions. The majority of these jets originates from parton-parton collisions mediated by strong interactions. Such processes have been calculated in the framework of perturbative QCD and the calculations have shown good agreement with the available data on jet production. The experimentally observed jets are a mixture of quark and gluon jets. Their relative frequency is expected to be a strong function of kinematical variables. Jet events with invariant masses small compared to the $p\bar{p}$ center-of-mass energy are dominated by subprocesses involving one or more gluons.

Jets are also produced by electromagnetic or weak subprocesses. The production cross sections for such processes are far smaller than for QCD processes. A global analysis of jet events can only statistically investigate such processes with very small signal-to-background ratios. The observation of hadronically decaying W/Z bosons recently reported by the UA2 collaboration [1] represents an example for such an analysis. Except for final state corrections the jets emerging from such processes are exclusively quark jets. This feature provides the only potential selection criterion available to experiments on an event-by-event basis. Indications for differences between quark and gluon jets have been observed in $p\bar{p}$ collisions [2]. So far no attempts have been made using such differences to enrich one species of jets with respect to the other.

In e^+e^- annihilation the situation is quite different. Here the production of gluon jets is an $O(\alpha_s)$ correction to the $e^+e^- \rightarrow q\bar{q}$ process. The gluon jets are therefore generally less energetic than the quark jets. Using simply this fact one is able to separate quarks from gluons with typically $\sim 65\%$ identification rate [3].

Recently, attempts have been made to use a Neural Network (NN) approach to separate quark jets from gluon jets in e^+e^- [4] with very promising results. Treating the problem as a general pattern recognition task, where the pattern corresponds to the four-momenta of the four leading particles, and using a feed forward NN, it is possible to increase the identification rate to $\sim 85\%$.

This paper describes our attempts to apply NN techniques also in the case of $p\bar{p}$ collisions.

After briefly describing the NN algorithm used, this paper presents two different attempts to analyze $p\bar{p}$ collisions with a NN technique. The first method is based on single jets which have been identified by a standard jet finding algorithm. This method tries to experimentally separate jets into quark and gluon jets on a jet-by-jet basis. The second method uses the information from the entire $p\bar{p}$ collision without

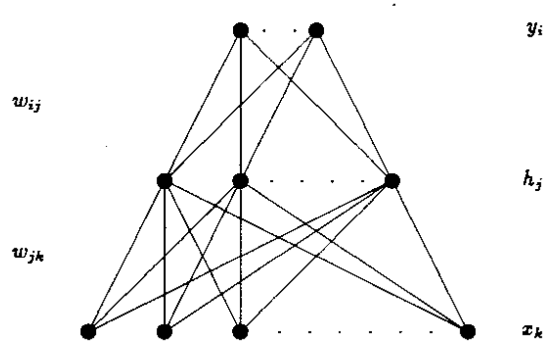

Figure 1: A feed-forward neural network with one layer of hidden units

requiring a jet identification first. Here the aim is to separate "W-type events" from QCD jet events. In this case the information due to the different initial state configuration of QCD processes and weak production of W bosons is used in addition to differences between the fragmentation of quark and gluon jets. This procedure is therefore specific to the case of hadron collisions.

2 The Neural Network Algorithm

In all our calculations we used a NN of the feed-forward hidden layered perceptron type as it is implemented in the program package JETNET [6]. Here we will only briefly describe this NN architecture. For a more detailed description we refer to [4] and references therein.

In a feed-forward NN, a set of neurons (or nodes) are arranged in a layered structure (see fig.1) with one input layer into which the information from the detector is fed, one (or several) hidden layers which allows the NN to build up its own internal representation of the input, and one output layer which represents the feature. Every neuron h_j in the hidden layer is connected to every neuron x_k in the input layer with connectivity weights w_{jk} and performs a weighted sum of the input signals and thresholds this sum with a sigmoid function g

$$h_j = g(\sum_k w_{jk} x_k / T) \tag{1}$$

Where the "temperature" T sets the slope of g. Correspondingly for the output neurons

$$y_i = g(\sum_j w_{ij} h_j / T) \tag{2}$$

Training the NN corresponds to changing the weights w_{jk} and w_{ij} such that a given pattern p of inputs $\vec{x}^p$ gives rise to an output $\vec{y}^p$ that equals the desired values $\vec{t}^p$. Here we used the back-propagation learning rule [5] where the least mean square error function

$$E = \frac{1}{2} \sum_p \sum_i (y_i^p - t_i^p)^2 \tag{3}$$

is minimized by changing the weights by gradient descent. For the weights between the hidden and output layer this gives us

$$\Delta\omega_{ij} = -\eta (y_i - t_i) g'(\sum_j w_{ij} h_j / T) h_j \tag{4}$$

and between the input and hidden layer we have

$$\Delta\omega_{jk} = -\eta \sum_i w_{ij} (y_i - t_i) g'(\sum_j w_{ij} h_j / T) g'(\sum_k w_{jk} x_k / T) x_k \tag{5}$$

here η is a learning strength parameter. This procedure is repeated until the NN has learned all patterns to a satisfactory level.

In the applications below, the input layer is fed with e.g. the transverse energy E_T measured in the cells of a calorimeter and the output layer, consisting of one neuron only, corresponds to the feature encoded as e.g. 0 for a gluon jet and 1 for a quark jet.

3 Separation of Single Quarks and Gluon Jets in $p\bar{p}$ Collisions

To see how well a NN can separate quark jets from gluon jets, a set of $p\bar{p}$ events at 630 GeV center of mass energy has been generated with the PYTHIA [8] Monte Carlo program. Only the $qq \to qq$ and $q\bar{q} \to q\bar{q}$ subprocesses were considered when generating quark jets and the $gg \to gg$ process when generating gluon jets. The energy of the fragmentation products was mapped into a calorimeter with a granularity of $\Delta\eta = 0.20$ in pseudorapidity and $\Delta\phi = 0.26$ in azimuth. This calorimeter had a complete coverage in ϕ and a pseudorapidity coverage of $|\eta| \leq 2$. The set-up corresponds to the UA2 calorimeter used in [1]. For each event, jets were reconstructed using the LUCELL algorithm provided by the JETSET [7] program. Jet transverse energies were collected in cones of radius 0.8 in η, ϕ space.

The produced jets were separated into one training and one test sample, each with equal amounts of quark and gluon jets.

The calorimeter information was presented to the NN in two different ways

(a) Take a matrix of 7×7 cells around the center of the jet and assign the E_T of each individual cell to one input node. This case corresponds to 49 input nodes.

(b) Take the E_T of the leading cell in the 7×7 matrix and assign it to the first node x_1. Assign the η and ϕ coordinates relative to the center of the jet to x_2 and x_3 respectively. Then take the second leading cell and assign its E_T, η and ϕ to x_4, x_5 and x_6 and so on for the first 15 cells. This corresponds to 45 input nodes.

For both cases (a) and (b) 10 hidden nodes were chosen. The effect of additional information presented to the NN has been studied by using the total E_T of the event and the total E_T in the forward and backward directions($1 \leq |\eta| \leq 2$) as additional input nodes. The effect of hadronic and electromagnetic shower development has been studied by smearing the energy content of each cell uniformly over the adjacent cells.

After adjusting the weights using the training algorithm and the training set, the NN was applied to the test set in order to check how well it has learned to recognize the features and how well it is able to generalize. The simplest way of doing this is to place a cut on the output node y_1 such that if the NN for a specific jet produces an output $y_1 > 0.5$ the jet is classified as a quark jet, and vice versa. The number of correctly classified jets is then a measure of the NN performance. In both cases (a) and (b) the NN was able to classify 70-72% of the jets correctly.

In most cases however, this is not a very interesting number. Usually the NN will be used as a trigger by separating out a signal (e.g. quark jets) from a background (e.g. gluon jets). Both, efficiency and signal-to-background ratio have to be optimized. The performance of the NN is best represented by a plot showing the signal-to-background ratio as a function of the signal efficiency.

In fig.2 such a plot is shown for the two cases (a) and (b). Strategy (b) gives far better results. This is because here the most significant calorimeter cell is always assigned to the same input node, whereas in case (a) the most significant node varies from pattern to pattern, making it more difficult for the NN to learn.

In fig.3 the same plot is shown for the (b) case where we have varied the amount of information presented to the NN.

1 Two extra input nodes, one for the total transverse energy in the event and one for the total forward and backward transverse energy summing over the pseudorapidity range $1 \leq |\eta| \leq 2$.

2 One extra node for the total transverse energy in the event.

3 The standard case (b) as in fig.2.

4 As in fig.2 but the contents of each cell is smeared out uniformly over the adjacent cells to provide a crude simulation of electromagnetic and hadronic shower spreads.

For large quark jet efficiencies curve 4 in fig.3 resembles the curve of case (a) in fig.2. This indicates that the NN in case (a) tries to average over adjacent cells to cope with the problem of significance mentioned above.

4 $W/Z \to$ 2 JETS

The NN described in the previous section has been designed to act on individual jets. The following study is an attempt to use the entire event information without explicit jet reconstruction to distinguish hadronic decays of vector bosons from the dominating QCD background.

When deciding on what to show the NN when training it to discriminate between different features it is important to think about what the NN is intended to be used for. In this case we want the NN to recognize hadronically decaying W/Z bosons such that it can be used to suppress the QCD background for jet spectroscopy. Any

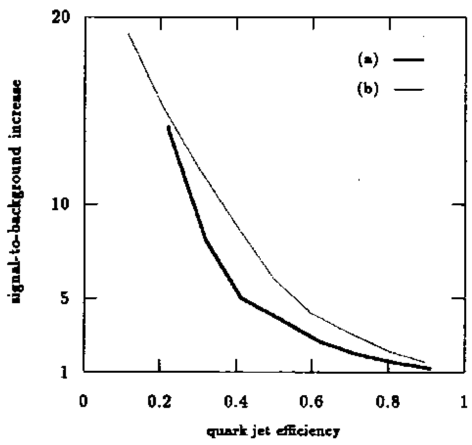

Figure 2: The increase in signal-to-background ratio vs. the efficiency for a NN used as a quark jet trigger for the two different strategies (a) and (b).

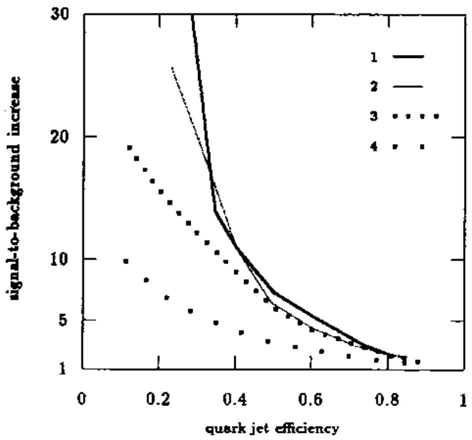

Figure 3: The increase in signal-to-background ratio vs. the efficiency for a NN used as a quark jet trigger for the (b) case with varying amount of information given to the NN as described in the text.

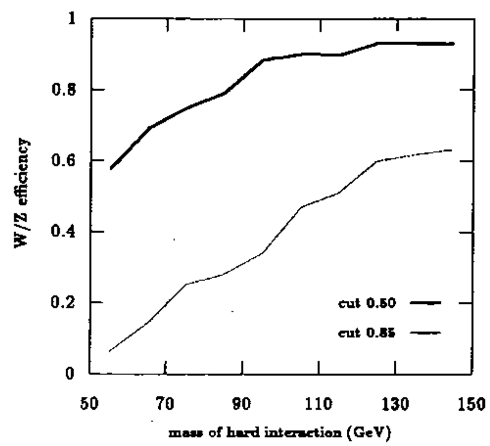

Figure 4: The efficiency for a NN used as a W trigger as a function of the mass of the hard interaction for two different cuts.

information about about the boson mass presented to the NN would be a trivial selection criterion and has to be avoided.

We generated W events, again using PYTHIA, with randomly chosen W masses between 50 and 150 GeV. We also generated an equal amount of QCD background events with a flat mass spectrum. In both cases at least two jets were required in the rapidity interval $|\eta| \leq 0.7$. For each event the 30 highest transverse energy cells were presented to the NN using a strategy similar to case (b) of the previous section. We used a NN with two hidden layers; 90 input nodes, 20 nodes in the first hidden layer and 4 in the second.

It turns out that even if the mass spectrum of the events in the training set was flat, the efficiency and background suppression for a given cut on the output node is not constant over the mass interval as is seen in figs.4 and 5. Efficiency and background suppression could be made constant by using a mass dependent cut. This has however not been done in the following analysis.

The NN has been tested on events with natural mass distributions. In fig.6 the invariant mass of the two largest jets of the QCD background is shown without and with the NN trigger using two different cuts on the output node. In fig.7 the same is

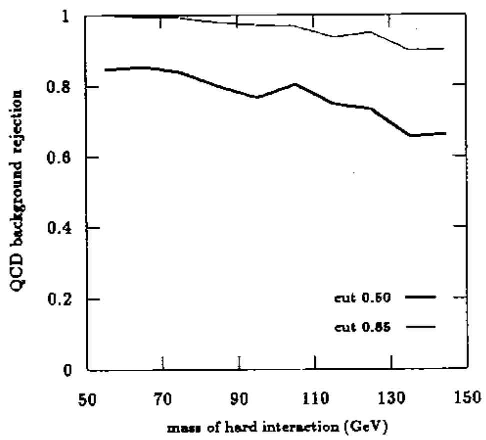

Figure 5: The background rejection for a NN used as a W trigger as a function of the mass of the hard interaction for two different cuts.

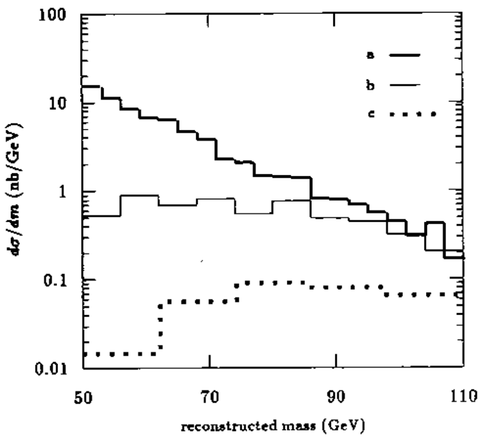

Figure 6: The two-jet invariant mass distribution for the QCD background (a) without trigger, (b) with NN trigger cut=0.50 and (c) cut=0.85.

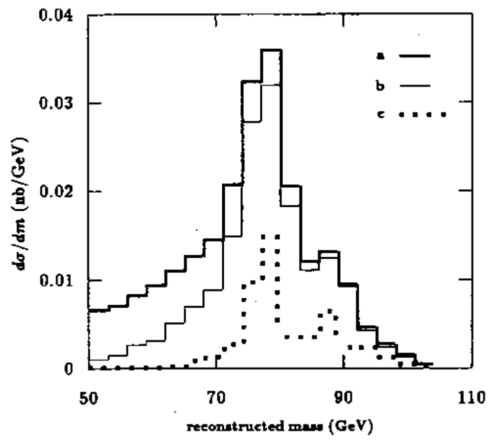

Figure 7: The two-jet invariant mass distribution for the W and Z signal (a) without trigger, (b) with NN trigger cut=0.50 and (c) cut=0.85.

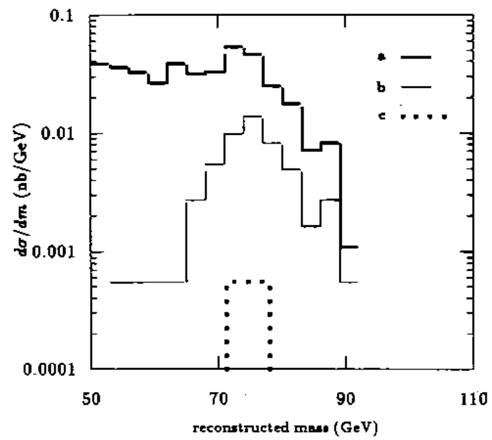

Figure 8: The two-jet invariant mass distribution for a gluonic resonance with mass = 80 GeV. Notations as in fig.7.

shown for the W and Z signal. Looking at the W and Z peaks, it appears that for the weaker cut the efficiency is almost 1, while for the stronger cut the signal is reduced a factor 2. The QCD background in this region is however reduced by factors 3 and 20-50 respectively (due to the large rejection the statistics is very low for the strong cut on the QCD background).

Note also that the W and Z peaks tend to get sharper when the NN trigger is applied. This is partly due to the mass dependence of the trigger efficiency. It also indicates that what the NN considers to be a typical W or Z is an event with two well collimated (quark) jets. If the jets have been broadened through QCD bremsstrahlung, hence giving a lower reconstructed mass, they are recognized as a gluon jets and the event is classified QCD background.

To support this interpretation we also generated an artificial gluonic resonance with the same mass as the W, decaying into two gluons. Applying the same cut on the output node as before, yields the two-jet mass distribution shown in fig.8. We see that for the lower cut the peak is significantly sharpened and reduced, while for the stronger cut it has almost vanished.

5 Summary

We have studied the possibility of using a NN for jet analysis in $p\bar{p}$ collisions. Attempts were made to identify experimentally observed jets as quark jets or gluon jets.

In a global event analysis the characteristic features of hadronically produced and decaying W bosons were used to distinguish them from QCD jet events. The originally very small signal-to-background ratio for this case could be increased by more than a factor 10 using a NN trigger. These methods provide exciting possibilities for the use of jets in the analysis of hadronic collisions. The results presented in this paper can be checked with existing data from the current hadron collider experiments.

It should be noted that parallel to the work of our group a study to find more simple ways of discriminating between quark and gluon jets [9] has been performed.

Using simple variables such as e.g. the minimum number of cells in a jet which contribute 90% of the jets E_T and the number of cells in the jet with $E_T > 1$GeV, it is possible to obtain efficiencies and signal-to-background ratios which are comparable with, and in some cases better than the NN approach.

Acknowledgment

We would like to thank Bryan Webber for useful discussions.

References

[1] UA2 Collaboration, J. Alitti et. al., A measurement of Two-Jet Decays of the W and Z Boson at the CERN $p\bar{p}$ Collider, CERN Preprint CERN-PPE/90-105.

[2] UA1 Collaboration, G. Arnison et. al., *Nucl. Phys.* **B276** (1986) 253.

[3] JADE Collaboration, W. Bartel et. al., *Z. Phys.* **C21** (1983) 37.

[4] L. Lönnblad, C. Peterson and T. Rögnvaldsson, *Phys. Rev. Lett.* **65** (1990) 1321, L. Lönnblad, C. Peterson and T. Rögnvaldsson, Using Neural Networks to Identify Jets, Lund Preprint LU TP 90-8 (to be published in Nucl. Phys. B).

[5] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning Internal Representations by Error Propagation", in D. E. Rumelhart and J. L. McClelland (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition (Vol. 1)*, MIT Press (1986).

[6] L. Lönnblad, C. Peterson and T. Rögnvaldsson, JETNET 1.0 program and manual.

[7] T. Sjöstrand, JETSET 7.3 program and manual. See B. Bambah et al., QCD Generators for LEP, CERN-TH.5466/89.

[8] H-U. Bengtsson, T. Sjöstrand, *Computer Phys. Comm.* **46** (1987)43.

[9] J. Pumplin, *contribution to these proceedings.*