b-Tagging Using Shape Variables in the Hadronic Decays of the Z^0

DELPHI Collaboration

P.Branchini[1], *M.Ciuchini*[1], *G.Cosmo*[2] *U.Daurù*[2],
A.DeAngelis[2], *P.DelGiudice*[1], *L.Lyons*[3], *A. Saccavini*[2], *F. Stringhetta*[2]

Abstract

The performance of b-taggers, based on Multidimensional Analysis applied to variables related to the shape of hadronic events, is analyzed on a sample of about 166,000 hadronic decays of the Z^0 collected by the DELPHI detector at LEP during 1991. The classification efficiency is computed from the data, by using tagging on hemispheres.

(Preliminary analysis on 166,000 events)

[1] Istituto Superiore di Sanità and INFN Sanità, Roma
[2] Dipartimento di Fisica dell'Universita' di Udine and INFN Trieste
[3] Department of Physics, Oxford University

1 Introduction

The selection of as pure as possible subsamples of hadronic decays of the Z^0 in which the primary branching is into $b\bar{b}$ pairs is of primary interest in LEP physics. Methods so far explored to achieve this are based on:

1. the shape of the multihadronic events [1];
2. the presence of leptons with large transverse momentum with respect to the event axis [2];
3. the presence of tracks with large impact parameter with respect to the primary vertex [3].

Method 2. is limited by low efficiency, since the branching ratio of b into muons is of the order of 10%. Method 1. is generally limited by low efficiency, and by a large systematic error in estimating the performance. The systematic error has usually been estimated using simulations, and the results are sensitive to the models assumed for the description of the hadronization process.

In this paper, we discuss the possibility to extract a discriminant variable from many variables related to the shape of a multihadronic event, and to evaluate its performance in an almost simulation independent way, using a feed-forward Neural Network as a multidimensional classifier.

In the past three years many attempts have been done to use Neural Networks as a tool for classification of jet events according to the flavour of the parent quarks (for a review of the results recently obtained see Ref. [6] [7] [8]). These methods achieved performances which are competitive with standard analysis in terms of efficiency and purity; they share, however, a weakness in the dependence on Monte Carlo, that makes the estimation of systematic errors particularly tricky.

This motivated us to explore a strategy (see for example [5] for a previous application of this method) based on the independent tagging of suitably defined emispheres in which events are divided. In this way, an almost Monte Carlo independent tagging can be obtained, paying the price of a worsening of the performances.

We report here on preliminary results obtained in this direction.

2 Experimental Procedure and Event Selection

The sample of events used in the analysis was collected during 1991 by the DELPHI detector at the LEP e^+e^- collider, operating at centre of mass energies around the Z^0 peak.

A description of the apparatus can be found in Ref. [9]. Features of the apparatus relevant for the analysis of multi-hadronic final states (with emphasis on the detection of charged particles) are outlined in Ref. [10]. The present analysis relied on the information provided by the central tracking detectors.

The central tracking system of DELPHI covers the region between 25° and 155° in polar angle, θ, with reconstruction efficiency near 100%. The average momentum resolution for the charged particles in hadronic final states is in the range $\Delta p/p \simeq 0.001p$ to $0.01p$ (p in GeV/c), depending on which detectors are included in the track fit.

Charged particles were used in the analysis if they had:

(a) momentum larger than 0.1 GeV/c;

(b) measured track length in the TPC greater than 30 cm;

(c) θ between 25° and 155°;

(d) relative error on the measured momentum smaller than 100%.

Hadronic events were then selected by requiring that:

(α) the total energy of the charged particles in each hemisphere (θ above and below 90°) exceeded 3 GeV;

(β) the total energy of the charged particles exceeded 15 GeV;

(γ) there were at least 5 charged particles with momenta above 0.2 GeV/c.

In the calculation of the energies, all charged particles have been assumed to have the pion mass.

A total of 166K events satisfied these cuts. Events due to beam-gas scattering and to $\gamma\gamma$ interactions have been estimated to be less than 0.1% of the sample; background from $\tau^+\tau^-$ events was calculated to be less than 0.2%.

The influence of the detector on the analysis was studied with the simulation program DELSIM [11]. Events were generated with the JETSET 7.3 Parton Shower Monte Carlo program [12] (JETSET PS in the following) with parameters tuned as in [13]. The particles were followed through the detailed geometry of DELPHI giving simulated digitizations in each detector. These data were processed with the same reconstruction and analysis programs as the real data.

3 The Double Tagging Method

The major drawback of multidimensional methods is that the dependence on the simulation introduces on the physical observable a systematic error which is harder to estimate compared to the single variable case.

In order to overcome this problem, a double tagging method was used in the present analysis.

Hadronic events were split into two hemispheres according to the plane perpendicular to the thrust axis, then only hemisphere-defined variables were considered. Let ϵ_b be the probability of tagging a hemisphere in a b event, and ϵ_l the same probability in u, d, s, c events. In the hypothesis in which the two hemispheres are statistically uncorrelated the following equations hold:

$$f_1 = \epsilon_b R_b + \epsilon_l (1 - R_b) \qquad (1)$$

$$f_2 = \epsilon_b^2 R_b + \epsilon_l^2 (1 - R_b) \qquad (2)$$

where f_1 is the fraction of tagged hemispheres and f_2 is the fraction of events in which both hemispheres are tagged. The quantities ϵ_b and ϵ_l can be determined from the equations above, by assuming $R_b = 0.217$ as predicted by the Standard Model.

It would be possibile to estimate R_b using a third equation coming from a lepton tagged sample but this possibility will not be considered here.

The correlations between hemispheres, and possible differences in efficiency between the u, d, c, s quarks, modify equation (2) in such a way that

$$f_2 = \epsilon_b^2 (1 + c_b) R_b + \epsilon_l^2 (1 + c_l)(1 - R_b) \qquad (2')$$

The coefficients c_b and c_l have to be determined by simulation:

$$c_j = \frac{f_{j,2}}{f_{j,1}^2} - 1 \qquad (3)$$

where $j = b$ or l.

The purity π of a selected sample of b-tagged events (i.e., of events in which both hemispheres are tagged as b) can be obtained from ϵ_b and ϵ_l through

$$\pi = \frac{\epsilon_b^2(1+c_b)R_b}{\epsilon_b^2(1+c_b)R_b + \epsilon_l^2(1+c_l)(1-R_b)} . \quad (4)$$

4 Variables Used for the Classification

Among many variables which have been considered for feeding the network, we have chosen ten variables on the basis of their discriminating power according to the F test. Impact parameter has been excluded because our intention is to combine the forthcoming results with the one obtained Ref [4] using impact parameter variables.

This procedure resulted in the following list of input variables:

1. The sphericity S^β calculated after a boost $\beta = 0.96$ along the thrust axis.
2. The aplanarity A^β calculated after a boost $\beta = 0.96$ along the thrust axis.
3. The sum of the momenta parallel to the event axis, after the boost.
4. The sum of the products of the momentum components parallel to the event axis, times the perpendicular components.
5. The sum of the p_T^2.
6. The longitudinal momentum of the most energetic particle.
7. The invariant mass M_{1234} of the four most energetic particles.
8. The directed sphericity S_{1234} of the four most energetic particles. 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 thrust direction in the laboratory frame.
9. The directed sphericity S_{123}.
10. The directed sphericity S_{124}.

The values of their F-test from the simulation, where

$$F = \frac{|<x_b> - <x_{udsc}>|}{\sqrt{\sigma_b^2 + \sigma_{udsc}^2}} \quad (5)$$

are listed in Table 1.

5 The Neural Network

We adopted a feed-forward Neural Network architecture trained with Backpropagation. Both the architecture and the learning procedure have become by now quite well known, so we

Variable	F
2	0.259
1	0.236
8	0.215
6	0.209
10	0.207
9	0.128
5	0.071
4	0.061
7	0.051
3	0.013

Table 1: *Value of the F-test for the variables used as an input to the multidimensional classifiers.*

refer to the literature for details (see [14]), just giving here a brief sketch of them and a list of the choices made for the parameters.

The Neural Network we used has 10 input variables, strictly coming from the shape of the event.

Each node in the input layer was associated with one of the input variables listed above. There were 2 hidden layers of 10 and 8 nodes and one output node. Each unit gives as output a sigmoid function of the weighted sum of its inputs.

The training procedure performs a gradient descent in the space of the weights with respect to a quadratic cost function quantifying the discrepancy between the value obtained at the output for each training event and the one conventionally fixed as the target value for the corresponding class (in our case 1 for b and -1 for udsc). The process is controlled by the "learning strength parameter" η and the "momentum" α [14]. Each updating step in the space of weights, computed by gradient descent, is multiplied by η and added to the previous step, multiplied by α.

The weights were updated every 10 events, chosen at random from the two classes $u\bar{u}$+$d\bar{d}$+$s\bar{s}$ +$c\bar{c}$ and $b\bar{b}$, in such a way that, on average, there was an equal number of events from each of the two classes.

Changing the parameters η and α during the training is convenient in order to allow for a fast movement in the space of weights in the early stage of training, and to obtain a controlled approach to the minimum at the later stage. For this reason, the learning and momentum parameters were decreased and increased respectively after every 3,000 updates (an "epoch") according to the rule:

$$\eta_t = \eta_{t-1} \times (\eta_{min}/\eta_{t-1})^{k_\eta}$$

$$\alpha_t = \alpha_{t-1} \times (\alpha_{max}/\alpha_{t-1})^{k_\alpha},$$

where η_{min} and α_{max} are the minimum (maximum) allowed values for the parameters, and subscript t $(t-1)$ refers to the epoch number. Exponents k_η and k_α were set to 0.05 and 0.14, respectively. Given the finite value of the weight change, the gradient descent might lead to an occasional increase of the error value. In this case, the parameters were reset to their initial values.

The architecture of the network is summarized in Table 2, together with the parameters used in the training phase.

Nodes in the hidden layers	10,8
α (training)	0.4 - 0.9
η (training)	0.05 - 0.0001

Table 2: *Characteristics of the NN.*

After training, the network was exposed to a new 'test' set of events; the behaviour of the error function with respect to this new set shows the error the network makes in generalizing to new data.

It is relevant to monitor the generalization error for trainings of increasing lenghts; when the generalization error starts to increase while the training error is still decreasing, the NN has 'overlearned' the training sample and its ability to generalize is degraded.

The system was trained on a set of 20,000 simulated events, generated with the JETSET Parton Shower Monte Carlo model (JETSET PS), where each event is described using variables related to a single hemisphere. By monitoring the behaviour of the generalization error, the training was stopped after 400,000 updates. After the network has been trained, its performance can be assessed in terms of signal efficiency ε_s (number of events correctly classified as belonging to a given class over the total number of events in that class), and purity p (number of correctly classified events in a given class over all the events classified as belonging to that class). The test sample consisted of about 50,000 simulated events, generated by using JETSET PS and 50,000 real data taken during 1991 using the DELPHI detector.

The distribution of the output of the network is plotted in Figure 1(a) for b and $udsc$ events. In Figure 1(b), the distribution of the network output on single hemispheres is shown for simulated and real data. In Figure 1(c), the fraction of events for which the network output is larger than a given cut in both hemispheres is plotted for real and simulated data.

The calculation of the coefficients defined in Eq. (3) gave the results plotted in Figure 2. Each event is classified testing the Network separately on both the hemispheres and considering an event as $b\bar{b}$ only if it is classified as b on both the hemispheres. The number of events generated in each class corresponded to the Z^0 hadronic branching fractions in the Standard Model.

The curve giving purity of the selected sample versus efficiency for the selection of b quark pairs is given in Figure 3. The shape plotted as black markers refers to real data, while the solid line is for the simulation. Purity and Efficiency are computed using the solutions of the equations (1) and (2') described in Section 3. The error bars are computed allowing a 100% variation of the c_j.

6 Summary and Conclusions

Using the independent information coming from suitably defined emispheres in jet events proves to be a viable strategy to obtain a multidimensional tagging method based on Neural Networks, which is less sensitive to Monte Carlo assumptions than previously achievable. Further analysis is required to better quantify the influence of the correlation coefficients on the classification in terms of efficiency and purity; this work is presently in progress. As one of the developments planned for the near future, we also expect to obtain better performances by the combined use of this method together with an independent, single particle tagging based on impact parameter ([4]).

Acknowledgements

We are greatly indebted to our technical collaborators and to the funding agencies for their support in building and operating the DELPHI detector, and to the members of the CERN-SL Division for the excellent performance of the LEP collider. Part of this work has been supported by INFN through the ANNETTHE initiative.

References

[1] W. Braunschweig et al., *Zeit. Phys.* **C44** (1989) 1, P. Abreau et al., *Phys. Lett.* **B281** (1992) 383

[2] D. Decamp et al., *Phys. Lett.* **B259** (1990) 236, P. Abreau et al., *Zeit. Phys.* **C56** (1992) 47, B. Adeva et al., *Phys. Lett.* **B261** (1991) 177, P.D. Acton et al., *Zeit. Phys.* **C55** (1992) 191

[3] The ALEPH Collaboration, CERN-PPE/93-108, June 1993, T. Baroncelli et al, DELPHI Note 93-93

[4] T. Baroncelli et al, DELPHI Note 93-93

[5] The ALEPH Collaboration, CERN-PPE/93-113, July 1993

[6] "New computing techniques in physics research II", D. Perret-Gallix ed., World Scientific 1992

[7] "Neural networks: from biology to high energy physics", supplementary issue of the *International Journal of Neural Systems*, World Scientific 1993

[8] "Computing in high energy physics", CERN 92-07, Geneva 1992

[9] P. Aarnio et al. (DELPHI), *Nucl. Instr. and Meth.* **A303** (1991) 233.

[10] P. Aarnio et al. (DELPHI), *Phys. Lett.* **B240** (1990) 271.

[11] *DELSIM User's Guide*, DELPHI 89-15 PROG 130, CERN, February 1989. *DELSIM Reference Manual*, DELPHI 89-68 PROG 143, CERN, September 1989.

[12] T. Sjöstrand, Comp. Phys. Comm. **27** (1982) 243, ibid. **28** (1983) 229; T. Sjöstrand, *PYTHIA 5.6 and JETSET 7.3*, CERN-TH.6488/92, September 1992.

[13] W. de Boer et al., IEKP-KA/91-07, Karlsruhe, June 1991.

[14] J. Hertz, A. Krogh, R.G. Palmer, "Introduction to the theory of neural computation" Addison-Wesley 1991

[15] Fisher.