Event Discrimination at LHC Exploiting Invariant Mass Topology[#]

R. Odorico*

University of Bologna, Physics Department
INFN, Sezione di Bologna
Via Irnerio 46, I-40126 Bologna, Italy
E-mail: odorico@bo.infn.it

ABSTRACT

A Neural Network approach for the discrimination of LHC events according to their invariant-mass topology is presented. Results obtained by running the neural network over the LHC event samples of the EAST benchmark for $H \rightarrow t\bar{t} \rightarrow e3j$ are shown. The neural network, once implemented on a dedicated neural microprocessor, can be used as part of the on-line trigger.

SW 95 42

1. Introduction

The discrimination of events at LHC containing semileptonic or non-leptonic decays of Higgs particles or top quarks represents a formidable task because of the complexity of the involved signatures. In principle, the decay patterns one is looking for are rather selective. For the Higgs, for instance, the dominant decay mode is expected to be into a pair of top quarks, with each one of them decaying into a W plus a bottom quark, followed by the decay of the W into a pair of jets or of leptons and that of bottom into charm. At each stage there are invariant-mass constraints to satisfy, in addition to lepton signals and impact parameter signatures. The problem one has to face is to grasp the complex system of interrelationships and to convert it into an operating procedure for event discrimination.

Neural Networks represent a convenient way of capturing statistical correlations within a set of physical variables, which are characteristic of a certain class of events [1]. Their application to phenomenological problems in high energy physics (HEP) was proposed some time ago [2]. In recent years, their use for the discrimination of events of interest in the off-line analysis of experimental data has widened (see, e.g., the proceedings of AIHENP'95 [3]). Using dedicated neural network chips, the technique has also been employed in the realization of non-leptonic triggers in experiments looking for heavy flavor production [4-8].

[#] Postscript file: <http://www.bo.infn.it/preprint/odorico.html>

* e-mail: odorico@bo.infn.it

Although neural networks represent a powerful additional tool for increasing the signal/background ratio, one should not indulge in the idea of using them as general-purpose black-boxes needing little care for the specification of their input and requiring little innovation in their architecture. There are intrinsic limitations they face, which become especially severe when implementing them on dedicated hardware. In addition, in HEP applications one easily meets problems not encountered in other fields. That is true in particular for the problem at hand: finding a way to discriminate events by means of their invariant-mass topology. Besides the architecture of the neural net, one must care for the choice of the input variables: they must be effective, and on-line calculable if meant for a trigger.

Aim of this note is to show that event discrimination based on the invariant-mass topology of multistep decays is amenable to a neural network approach, for a suitable choice of the input variables and a specific architecture of the net. The quantitative results presented refer to the LHC event samples of the EAST benchmark for $H \rightarrow t\bar{t} \rightarrow e3j$ and the associated background. These are not the most convenient data for the sake of this study, which is dominantly meant for non-leptonic decays, but they present the advantage of belonging to an established benchmark and thus offer a transparent way of evaluating performance.

Section 2 outlines the main feature of feed-forward neural networks. Section 3 presents the proposed neural network sensitive to the invariant-mass topology of events. Section 4 introduces the EAST benchmark. Section 5 reports on the performance of the neural network on the EAST benchmark. Section 6 contains the conclusions.

2. Feed-Forward Neural Networks

A feed-forward neural net is a mathematical construct consisting of a succession of layers. For each one of them, the standard feed-forward propagation expression holds:

$$o_i = \lambda_i \sum_j w_{ij} y_j + \theta_i \quad (1)$$

where y_i are the variables in input to the layer, w_{ij} are the weights, θ_i are the thresholds, λ_i are the scalar multipliers and o_i are the layer outputs. The latter can be sent in input to the next layer, after having been singularly filtered through a transfer function $T(x)$. Popular choices for $T(x)$ are $\tanh(x)$ or the sigmoid function $(1+\tanh(x))/2$. Optimization for the neural net parameters can be obtained using gradient minimization and the chain derivation rule applied to, e.g., the sum of the Euclidean distances between actual and desired outputs over a set of "training" input patterns (back-propagation algorithm).

The transfer function $T(x)$ must not be linear, otherwise the multilayer structure would be equivalent to a single layer net. For the rest, it is arbitrary. For training purposes, though, it is convenient for it to have easily computable derivatives, e.g.: $d(\tanh(x))/dx = 1 - \tanh(x)^2$.

A feed-forward neural net can be regarded as an approximation to a multivariate vector function, which is determined by a so called "learning by examples" procedure. It is a tool which can be used in a large variety of applications. In HEP it has been mainly used, so far, for event discrimination.

3. Neural Network Sensitive to Invariant-Mass Topology

An intrinsic limitation of standard feed-forward neural nets is that they are sensitive only to linear combinations of the input variables. In principle, that represents a problem if we need to design a neural net sensitive to the specifics of a multistep decay topology, where

consideration of invariant masses for more than two and in general an arbitrary number of particles/jets becomes important.

That obstacle can be overcome and we can obtain a neural net probing the entire topology of the event by taking as input variables the *squares* of invariant masses of particle/jet pairs.

Let us consider the system of four-momenta $p_1, p_2, \dots, p_N$ and let us denote by $m^2(1,2,\dots,N)$ its total invariant mass. By setting invariant masses of single four-momenta equal to zero, one can write:

$$m^2(1,2,\dots,N) = (p_1 + p_2 + \dots + p_N)^2 = \frac{1}{2} \sum_{i,j=1,N} m^2(i,j) \quad (2)$$

Thus the neural net can be made sensitive to any invariant mass combination which is found to be relevant during training.

In order to make the neural net respond to topologies for invariant masses associated to given decay patterns (like e.g. $H \rightarrow \tilde{t}\tilde{t} \rightarrow e3j$) one can conveniently take a transfer function not of the usual sigmoid type but rather with a Breit-Wigner shape:

$$T(x) = \frac{1}{1 + x^2} \quad (3)$$

Transfer functions of thy type have been used in non-HEP applications (see e.g. [9]). The Breit-Wigner width is determined by the training for weights and scalar multipliers, while its position is determined by the associated threshold value.

For the second layer (and possibly higher ones) there are no specific features to be taken into account. One can also consider substituting this part with some other algorithm, or integrate it with other neural networks sensitive to other physical variables (e.g. [10]).

The calculation of the invariant mass squares of particle/jet pairs can be handled fast by a matrix processor chip, like the MA16 of Siemens [11-13]. Another MA16 can implement the neural network. The four-momentum components can be computed by look-up-tables once the particle/jet momentum and angles are available. Thus, the entire neural net and the associated preprocessing of input variables can be implemented for on-line triggering.

4. EAST Benchmark Data Sets

The model for the LHC level-2 trigger (L2) emerging from the work of the EAST study group consists of three phases [14]:

i) *RoI collection*. A Region-of-Interest (RoI) is a spatially limited area ('3D-road') in which the level-1 trigger (L1) has identified candidates for phenomena to be triggered on (electrons, photons, muons, jets). The raw data pertaining to RoIs have to be selected by some mechanism, termed RoI collection, guided by a L1 device (i.e. outside the L2 data stream).

ii) *RoI feature extraction*. That amounts to the extraction from the raw data of each RoI of physically significant features, like e.g. cluster or track parameters.

iii) *Global decision*. That is divided in two subphases:

iii-a) *RoI task*: combination of all subdetectors that relate to the same physics object, i.e. to the same RoI, so as to get, e.g., probabilities for particle identities by some algorithm, which may consist of a neural network.

iii-b) *Event task*: combination of information from all RoIs so as to achieve a global decision by some algorithm.

EAST has formulated a benchmark to evaluate proposed solutions for the Event task [15]. The benchmark is focused on Higgs leptonic and semi-leptonic decays, and does not include other physics channels, like e.g. hadronic production of the top quark or non-leptonic decays, which are more sensitive to invariant masses. The benchmark is based on RoI input data consisting of features extracted from four types of detector: calorimeter, transition radiation detector, preshower counter, and muon chambers. Generated events are filtered through the L1 trigger model, with full simulation based on the present detector models. The data are fully described in [16]. From them, a neural network based RoI task algorithm computes angles and energies of particles/clusters and particle/jet identification probabilities for each RoI in the event. The associated output file is given in input to the Event task. In its present form, the file includes 500 events for each one of the six Higgs decay channel considered, plus 500 background events containing QCD jets. The results of the current EAST model for the Event task algorithm are [15]:

	Accepted Signal %	Accepted Background %
1. $H \rightarrow ZZ \rightarrow 4e$	100	0
2. $H \rightarrow ZZ \rightarrow 4\mu$	99.5	0
3. $H \rightarrow ZZ \rightarrow 2e2\mu$	99.6	0
4. $H \rightarrow ZZ \rightarrow 2e2j$	96	0
5. $H \rightarrow t\bar{t} \rightarrow e3j$	94.6	5.6
6. $H \rightarrow t\bar{t} \rightarrow \mu3j$	100	0.4

Triggers for channels proceedings via ZZ use conventional cuts on particle/jet probabilities and invariant masses for pairs of particles/jets. Those for $t\bar{t}$ channels use only particle/jet probabilities.

5. Neural Network Performance on the EAST Benchmark

In order to have a first realistic indication on the performance of an invariant-mass sensitive neural net, a statistical study has been made using the EAST benchmark data. These data do not represent the most appropriate testbed for this kind of study: for the physics channels considered most of the selection can be handily done by exploiting particle/jet probabilities. Nevertheless, we have picked up the $H \rightarrow t\bar{t} \rightarrow e3j$ (i.e. electron + 3 jets) and the background data sets and trained a neural net to sort them out. Choice of this channel is due to the fact that there is still some, although small, room for improvement and to the relative complexity of the signature from an invariant-mass point of view. Fig. 1 shows the distribution of the two samples in the maximum electron probability p_e found in the particles/jets of an

event. From it one can see that there is a long tail in the background sample reaching out to $p_e = 1$.

The aim of the statistical study has been to see whether, by supplementing the p_e information with the output of an invariant-mass sensitive neural net, one can clean up the background in the $p_e \approx 1$ region. Simplicity of interpretation and of hardware implementation have been adopted as the guiding criteria in designing the net. The net has been made totally blind to p_e information, although ordering of the particles/jets according to their p_e could have improved performance. The net has 1 output, 3 hidden nodes, 28 input variables representing all possible two particle/jet invariant-mass squares among a maximum of 8 particles/jets (events with more than that represent less than 4% of the total: they have been dropped in this study), and Breit-Wigner shaped transfer functions.

Fig. 2 shows the distributions of the two event samples in the neural net output. Taken by itself it is not impressive, but things change when looking at Fig. 3 showing the scatter plots in the neural net output versus p_e . One can see now how in the top-right quadrant, i.e. for large neural output and large p_e , the background is eliminated (within the available statistics). With a large enough statistics for the background, one could likely improve performance by training the net on large p_e events only, especially to increase acceptance for the signal.

As to the response time with a MA16 implementation of the neural net, for a maximum of 8 particles/jets and with the above 2-layer net one would get the response in $\sim 7.5 \mu\text{s}$. That would be reduced to $\sim 4.5 \mu\text{s}$ when utilizing only the first layer and leaving the output from the corresponding nodes to be handled, together with other particle/jet information, by a separate module. To that one should add the preprocessing time needed to calculate invariant mass squares of particle/jet pairs. Doing it by a MA16 would require an additional $\sim 1.6 \mu\text{s}$ if using a pipelined transfer mode.

6. Conclusions

A neural network solution has been presented for the problem of exploiting the invariant-mass pattern exhibited by the multistep decay of heavy objects, like the Higgs meson or the top quark, for event discrimination at LHC energies.

Key ingredients of the neural network are:

- i) usage of invariant-mass *squares* of particle/jet *pairs* as input variables;
- ii) adoption of a Breit-Wigner shaped transfer function.

Picking up the $H \rightarrow t\bar{t} \rightarrow e3j$ channel of the EAST benchmark as a case study (although it is not the best suited for this sake, a non-leptonic channel being the appropriate one), it has been verified that a neural network of this type substantially improves performance in event discrimination when combined with the standard electron probability information.

A neural network of this type, including the necessary input variable preprocessing, can be implemented on existing hardware to become part of a level 2 trigger.

Acknowledgements

I would like to thank the EAST Collaboration for making available the EAST benchmark data sets and other relevant documentation concerning its LHC trigger studies, and its chairman Rudi Bock for discussions.

References

- [1] 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), p. 318.
- [2] R. Odorico, Proc. of the 1982 DPF Summer Study on Elementary Particle Physics and Future Facilities, Snowmass, Colorado, 1982, p. 478; R. Odorico, Phys. Lett. 120B (1983) 219; G. Ballocci and R. Odorico, Nucl. Phys. B229 (1983) 1.
- [3] Proc. 4-th Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, 1995, Pisa, Italy, Ed. B. Denby and D. Perret-Galix (World Scientific, Singapore).
- [4] C. Baldanza et al. (ANNETTHE Collaboration), in Proc. 3rd Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, Ed. K.-H. Becks and D. Perret-Galix (World Scientific, Singapore, 1990), p. 391.
- [5] C. Baldanza et al. (ANNETTHE Collaboration), University of Bologna report DFUB 94/16 (1994), Nucl. Instr. and Meth. A, in press.
- [6] C. Baldanza et al., Proc. 4-th Int. Workshop on Software Engineering, Artificial Intelligence and Expert Systems for High Energy and Nuclear Physics, 1995, Pisa, Italy.
- [7] B. Denby et al., Nucl. Instr. and Meth. A335 (1993) 296.
- [8] B. Denby et al., Nucl. Instr. and Meth. A356 (1995) 485.
- [9] B.G. Giraud et al., Saclay report T94/047 (1994), Neural Networks in press.
- [10] P. Chiappetta et al., Phys. Lett. B322 (1994) 219.
- [11] U. Ramacher et al., in Proc. Int. Conf. on Systolic Arrays, Killarney, Ireland, Prentice Hall (1989), p. 277.
- [12] U. Ramacher, J. Beichter and N. Bruels, MA16 -- Programmable VLSI Array Processor for Neuronal Networks and Matrix-based Signal Processing, User Description, Version 1.4, Siemens AG, ZFE T ME2, 81730 Munich, Germany (available from N. Bruels, E-mail: nico@dali9.zfe.siemens.de).
- [13] N. Bruels, MA16 -- Development Software, User Manual, Version 1.0, Siemens AG, ZFE T ME2, 81730 Munich, Germany.
- [14] R.K. Bock, W. Krischer, 4-year Status Report to the DRDC: Embedded Architectures for Second-Level Triggering (EAST), CERN/EAST Note 94-23 (1994).
- [15] R.K. Bock, J. Carter, I.C. Legrand, Modelling of L2 Global Decision Structures, CERN/EAST note 93-03 (1993).
- [16] R.K. Bock, J. Carter, I.C. Legrand, J. Varela, Test Data for the Global Second Level Trigger, CERN/EAST Note 93-01 (1993).

Figure Captions

Fig. 1 - Distributions of the two event samples in the electron probability p_e .

Fig. 2 - Distributions of the two event samples in the neural-net output.

Fig. 3 - Scatter plots for the two event samples in the neural-net output versus p_e .

Fig. 1

Fig. 2

Fig. 3