1° Congresso Brasileiro de Redes Neurais

Escola Federal de Engenharia de Itajuba
Itajuba. 24 a 27 de outubro de 1994

Neural Particle Discriminator Based on Principal Components Analysis

J.M. Seixas , L.P. Caloba and B. Kastrup
COPPE/EE/UFRJ, C.P. 68504, Rio de Janeiro 21945, Brazil

Abstract

An electron/jets discriminator for high energy physics experiments is developed using neural networks. A principal components analysis is performed in order to reduce dramatically the input space dimension. so that system implementation becomes simpler. It is shown that using simulated calorimeter data, 95% electron efficiency is obtained with 9.5% of jets being misidentified, when input data is projected onto the subspace spanned by six principal components.

Introduction

The LHC (Large Hadron Collider) project [1] is being designed at CERN (Switzerland). It is planned to be operational by the beginning of the next century and it aims to bring new highlights on the study of the fundamental structure of the matter. This is to be achieved by colliding high-energy particles and analyzing the obtained reaction products by means of particle detectors placed all around the collision interaction point. The collision frequency for LHC is expected to be 40 MHz.

Among detectors. calorimeters became very important for collider experiments in the last years [2]. Particles interacting with these detectors deposit entirely their energy and the depositing process is such that calorimeters can also be used to identify the incoming particles. As the signals from the calorimeter can be quite fast, real-time event selection can be achieved, which becomes a desirable feature in the high event rate environment of LHC.

The event selection is performed by complex triggering systems and the operation of such systems is split into successive levels. The first-level trigger (L1) receives all generated events and using mainly the calorimeter information retrieves those events that seem relevant for the physics the experiment is interested on. For LHC, detailed physics simulations show that a second-level trigger system is desired to achieve further event rate reduction after the first-level operation. For the second-level trigger, calorimeters can be combined with other fast detectors. The expected rate at the second-level trigger input is 100 kHz.

The second-level trigger operation is being conceived in two phases [3]. In the first phase, the main features of each detector are extracted. This is performed over the information sent by detectors in a region of interest (ROI) identified by the L1 system. Next, these features are combined by means of a global decision unit and the events are selected according to the processes they represent. A rejection factor of 100 is expected from the whole system.

In this paper we consider the feature extraction problem for calorimeters. The features to be extracted can be found by searching for the best calorimeter variables one could send to the input nodes of a neural network discriminator capable to perform electron/jets separation based only on calorimeter information. For this purpose, Monte Carlo simulations have been used to generate events for the second-level operation [4]. A large sample of QCD jets and single electrons were generated and the interaction of those events with a fine-grained calorimeter was obtained by means of a 20x20 matrix of deposited energy in the ROI. On the sample data, a first-level trigger algorithm [5] was applied. In this manner, a total of 1057 jets and 1634 electrons were produced.

As the outermost cells of the calorimeter sample very few (or null) energy, it was possible to reduce the input space dimension through defining a subregion of 11x11 cells [6]. In order to define this subregion, the cell of maximum energy deposition was found in an event by event basis and the subregion was built around it. Figure 1 shows typical events.

Recent works had explored the neural network approach to perform feature extraction for calorimeters, examining different features [6, 7]. Results revealed that neural discriminators perform better than classical methods. We mention here the matrix and ring methods which were developed on those works and will be referred to in the next sections. The matrix method consists on feeding the input nodes of a neural network with the energy values of the 121 cells that form the subregion of interest. On the other hand, the ring method consists on building concentric rings around the cell of maximum energy deposition, so that this cell becomes the first ring (see Figure 2). The ring sums are obtained by adding up the energy of all cells that belong to a ring. The ring sums are then fed into a network to perform the desired discrimination. For 95% electron efficiency, the matrix and ring methods misclassified 7.2% and 9.7% of the jets, respectively. If one uses a weighting procedure for rings, which has the effect of boosting the energy of the

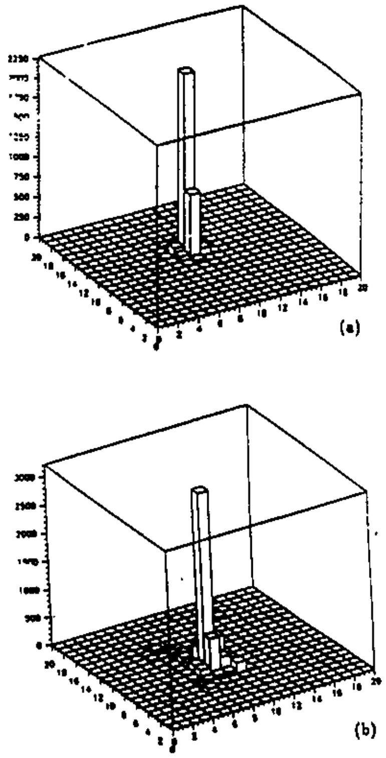

Figure 1. Typical electron (a) and jet (b) events.

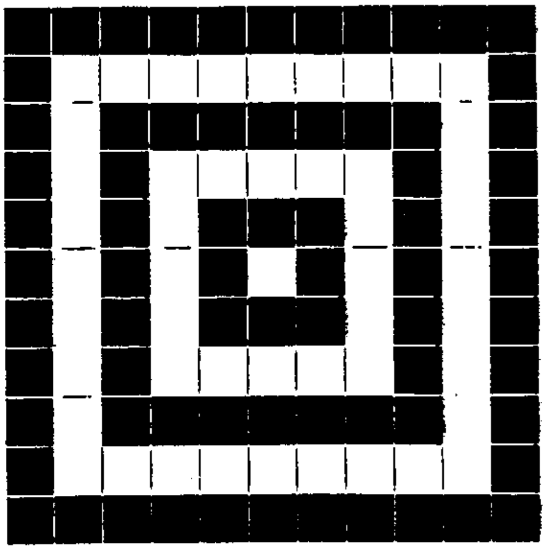

Figure 2. Building rings.

outermost rings, 8.5% of jets are misclassified for the same electron efficiency.

The next section describes the principal component extraction by means of a neural network. Then, it is shown that using at least 4 components, the proposed neural discriminator gets close to the performance obtained with ring sums.

Extracting the Principal Components

The principal components analysis (PCA) appears in many fields of application [8]. The aim is to find a set of M orthogonal vectors in input data space that account for as much as possible of the data's variance. Therefore, the differences in classes of events originally present in the data set can still be identified by projecting the data of the original N-dimensional input space onto the M-dimensional subspace spanned by these vectors. This would perform a dimensionability reduction with the preservation of the information spread around the full input space, as normally $M << N$.

In the case we are examining, the calorimeter information was reduced to

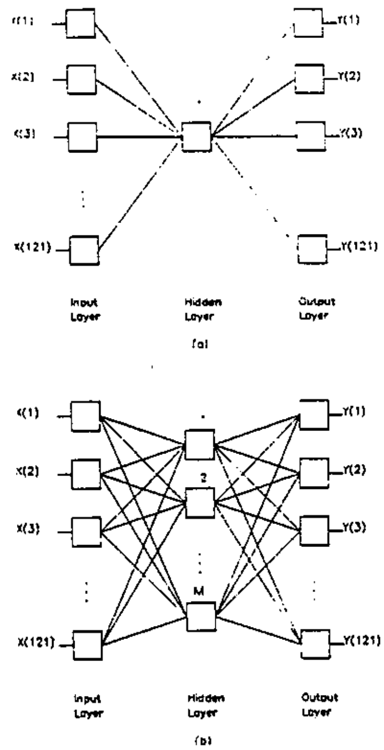

Figure 3. The network topology for extracting the first principal component (a) and the M-th component (b)

a set of 121 energy values that account to the energy deposited on each cell of the detector. Thus, one could think of further reducing this dimensionability by means of PCA, achieving a much more compact discriminator. The ring method mentioned in last section tries to do so using the concept of energy deposition in concentric rings. In fact, using the ring sums as a preprocessing of input data reduces the input space dimension for the discriminator network. As the 11x11 subregion of interest allows building up to six rings (see Figure 2), the input vector for ring analysis has only six components.

The way the principal components are extracted is the following. The first principal component is taken to be along the direction with the maximum variance. The second component is then constrained to lie in the subspace perpendicular to the first and it is taken along the direction of maximum variance. The other components are obtained by following the same procedure.

Neural networks architectures have been used to perform the principal components analysis [9]. Figure 3 shows a set of fully connected networks with one hidden layer used to do so. Each network is linear and has N inputs and N outputs. The first principal component is extracted by using one single unit in the hidden layer (Figure 3a) and training the network so that the output vector is as close as possible to the input vector. The hidden unit ends up revealing the first principal component in its weight vector. The other components are extracted in a similar way, by fixing the weights of the network used to extract the previous component, adding one more unit in the hidden layer and training the remaining network in the same way the first component was obtained (see Figure 3b). At the end of the training procedure, the network had extracted M principal components and the input data can be projected on the M-dimensional subspace spanned by the weight vectors in the hidden layer.

The mean square error (MSE) is the figure of merit in the training phase for deciding how many components one should extract for a given problem. After finding a certain number M of components and observing that MSE does not change significantly by incrementing the number of extracted components, one can consider that the set of M components extracted describes relatively well the distribution of events in the input space.

Results

In order to realize a neural electron/jets discriminator, one neural network was trained with projected events. As the number of principal components are normally much smaller than the dimension of the input space, PCA can be considered as a preprocessing method of input data that enables one to build a more compact discriminator system.

The neural networks used for extracting the principal components and for performing the electron/jets discriminator were simulated with the Jetnet 2.0 package [9]. The training file was built by dividing the total number of events by two. Testing was performed on events that do not belong to the training set. During the training phase of the network that realizes the discriminator, the target value was assumed to be -1 for electrons and 1 for jets. The hyperbolic tangent was the activation function for this network.

After extracting the fourth principal component, the discriminator start to be competitive. Using four components and respective projected data, 95% electron efficiency was achieved with 10.1% of jets misclassified as electrons. For this test, a 4-4-1 neural network performed the discriminator.

An interesting case to be analysed is the use of six principal components. In this case, one can compare the PCA preprocessing with the ring description. Extracting six components, using projected events and a 6-6-1 network topology to realize the discriminator, 9.5% of jets are wrongly classified for the same 95% electron efficiency. This result shows that increasing from four to six components has the consequence of improving performance of the neural discriminator, so that the resulting system reaches the same level of performance obtained by

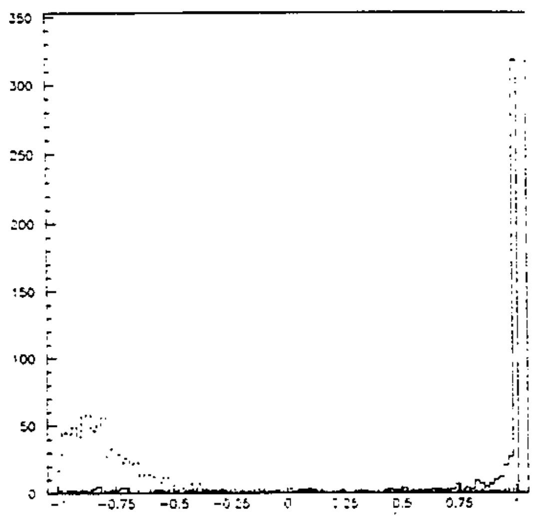

Figure 4. The network output using six principal components.

means of unweighted ring method. One should also keep in mind that the preprocessing needed for PCA is simpler than the one needed for building ring sums. Moreover. digital signal processor technology available nowadays allows the implementation of PCA preprocessing (which means performing inner products) integrated with the neural discriminator. Therefore. in terms of hardware implementation. neural discriminators based on PCA are rather attractive.

Figure 4 shows discriminator's output when six principal components are used. In terms of MSE figure of merit. increasing the number of components from four to six translates into a decrease of the mean square error by a factor of four.

Conclusions

A neural electron/jets discriminator based on principal component analysis was developed. It is based on performing a preprocessing of input data so that each event is projected on the subspace spanned by the principal components. As the number of components required to make the discriminator competitive with other discriminating methods is quite low (four components was the lowest limit found). the proposed discriminator allows a compact design. Moreover. both preprocessing and discrimination can be implemented using a single digital signal processor with the technology available nowadays.

The discriminator being proposed was tested on simulated calorimeter data for the second-level trigger operation on high-energy physics experiments. It was shown that using 6 (4) principal components. 95% electron efficiency is achieved with 9.5 (10.1) % of jets being misclassified as electrons. This performance is similar to the one obtained with a preprocessing that consists on forming ring sums from the original 11x11 matrix of energy deposited in the calorimeter. However. in terms of hardware implementation. the projection of events on the principal vectors is simpler than building ring sums. This makes the neural discriminator based on principal component analysis to be an option for the implementation of a second-level trigger at LHC.

Acknowledgements

We are thankful for the support that has been provided to this work by CNPq (Brazil) and CERN. We would like to thank the RD11 Collaboration for providing the simulated data sets. In particular. we express our best thanks to R.K. Bock. F. Block. W. Krischer. I.C. Legrand and J. Carter. all from CERN. for fruitful discussions concerning this work. and A. Ordacgy for helping in part of the analysis.

REFERENCES

1. The Atlas Collaboration. Letter of Intend for a General-Purpose pp Experiment at the Large Hadron Collider at CERN. CERN/LHCC/92-4, 1992.
2. R. Wigmans - Advances in Hadron Calorimetry. Rev. Nucl. Part. Sci. 41 (1991) 133
3. J. Badier et al. IEEE Trans. on Nuclear Science **40** (1993) 45.
4. G. Klyuchnikov et al. A Second-level Trigger Based on Calorimetry Only. **CERN-ATLAS/DAQ-007** (1992).
5. RD-27 Collaboration. **CERN-ATLAS/DAQ-005** (1992).
6. J.M. Seixas et al. A Second-Level Trigger System Based on Calorimeters and Using Neural Networks for Feature Extraction and Electron/Jets Discrimination. IV Int. Conf. on Calorimeters for High En. Phys. Elba. Italy. 1993.
7. J.M. Seixas et al. Neural Networks Applied to a Second-Level Trigger System. III Int. Workshop in Softw. Eng., Artificial Intell. and Expert Sys. for High-En. and Nucl. Phys. Oberammergau. Germany, 1993.
8. S. Haykin. Adaptive Filter Theory. Prentice-Hall (1991)
9. J. Hertz. A. Krogh and R.. G. Palmer. Introduction to the Theory of Neural Computation. Addison-Wesley (1991).
10. L. Lönnblad et al. Comput. Phys. Commun. **70** (1992) 167.