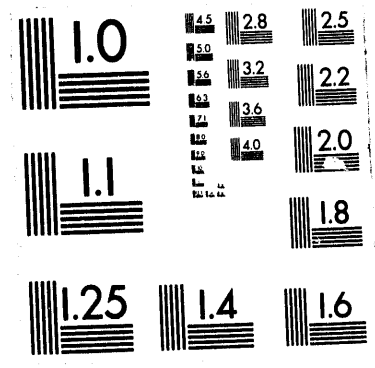
1.0
2.8
2.5
3.2
2.2
3.6
2.0
4.0
1.1
1.8
1.25
1.4
1.6

1 of 1

Conf-940699--1

"The submitted manuscript has been authored by a contractor of the U.S. Government under contract No. DE-AC05-84OR21400. Accordingly, the U.S. Government retains a nonexclusive, royalty-free license to publish or reproduce the published form of this contribution, or allow others to do so, for U.S. Government purposes."

Using Neural Networks as an Event Trigger in Elementary Particle Physics Experiments

Eric Neis*, Francis W. Starr#, and Thomas Handler
Department of Physics
University of Tennessee-Knoxville
Knoxville, TN 37996-1200

Tony Gabriel, Charles Glover, and Surender Saini
Oak Ridge National Laboratory
Oak Ridge, TN 37831

Abstract

Elementary particle physics experiments often have to deal with high data rates. In order to avoid having to write out all data that is occurring online processors, triggers, are used to cull out the uninteresting data. These triggers are based on some particular aspect of the physics being examined. At times these aspects are often equivalent to simple pattern recognition problems. The reliability of artificial neural networks(ANNs) in pattern recognition problems in many fields has been well demonstrated. We present here the results of a study on the feasibility of using ANNs as an online trigger for high energy physics experiments.

1. Introduction

Elementary particle accelerators are capable of producing vast numbers of particles in their energetic collisions. The events of interest are often imbedded in large numbers of uninteresting events. The proposed Large Hadron Collider(LHC) and the future replacement for the Superconducting Super Collider (SSC) will have event rates on the order of 10^8 events per second with the majority of these events being of no interest. As it is impossible to collect such large numbers of events with their complete data description, it is necessary to have a method to eliminate unwanted events. Often the events of interest exhibit some particular aspect that can be cast into a pattern recognition problem. At the SSC, the decay of the Higgs particle would evidence itself in the production of jets of highly energetic particles. These jets of particles would then be seen in the detector as clusters of energy deposition. Artificial neural networks are useful for cluster detection and thus are a prime candidate as a hardware trigger to filter the incoming data.

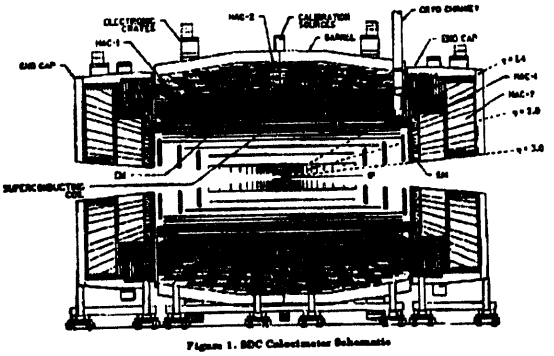

Figure 1. SDC Calorimeter Schematic

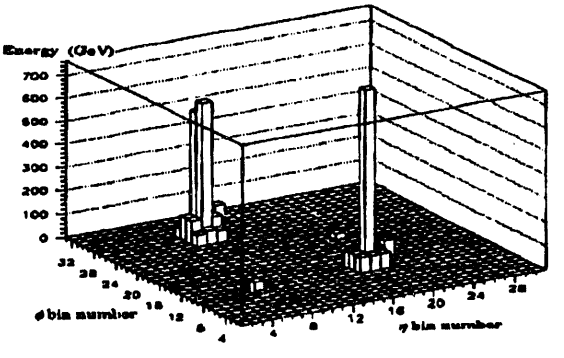

Figure 2 Typical Two Jet event

The deposition of the particles' energies would occur in the calorimeter segment of the detector. An example of the such a detector is shown in Figure 1. This detector would have been built by the Solenoidal Detector Collaboration at the SSC and is fully described in Reference 1. The calorimeter portion of the detector is segmented in azimuth and polar angle. The result of projecting these segments into a plane is pixel array of size 32 by 30. As

* University of Tennessee Science Alliance Summer Student 1992.
# University of Tennessee Science Alliance Summer Student 1993.

MASTER

*This research was sponsored by the Oak Ridge National Laboratory managed by Martin Marietta Energy Systems, Inc., under contract DE-AC05-840R21400 with the U.S. Department of Energy.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED
RECEIVED FEB 08 1994
OSTI

mentioned previously, events of interest would show up in this calorimeter projection as towers of large energy deposition.

As no actual data exists from the SSC, Monte Carlo events were generated using the program ISAJET [2]. Two classes of events were generated: 1) Minimum-bias events, i.e. events that are effectively background, and 2) Two jet events, i.e. events that would most closely resemble the decay of the Higgs particle. An example of a two jet event is shown in Figure 2. Details of the event generation procedures can be found in Reference 3. The generated events were then projected into the calorimeter and the energy depositions recorded. Events of both types were used both in the training and testing portion of this study.

II. Artificial Neural Network Development

A three layer fully connected artificial neural network was used in this study. The input layer consisted of 960 nodes, as dictated by the 30 by 32 calorimeter configuration. The networks tested varied in hidden layer size from as few as 10 hidden layer nodes to as many as 1000 hidden layer nodes. Only one output layer node was required as the output was either positive(two jet event) or negative(minimum-bias event). All network training and testing was done using the NeuralWorks Professional II/Plus software package [4]. 1500 events, 750 of each type, were used for training, while 2000 events, 1000 of each type, were used for testing the neural network.

The optimal size of the hidden layer was determined by considering its proficiency in identifying two jet events, since those are the events that would be kept for further analysis. Table I summarizes the average test set performance of the networks tested for five attempts at training.

TABLE 1: Determining Network Size

	Percent Correct	
Hidden Layer Nodes	Minimum Bias	Two Jet
10	95.7%	94.9%
20	98.1%	94.0%
30	98.3%	89.4%
40	99.5%	88.6%
50	97.3%	92.2%
60	98.8%	90.3%
70	97.2%	87.5%
80	96.4%	92.3%
90	99.5%	86.7%
100	98.4%	91.2%
1000	99.5%	78.6%

It is apparent from this table that all the networks were capable of identifying minimum bias events quite well. It is also clear that the smaller networks were better at identifying two jet events. This can be attributed to the fact that larger networks tend to memorize data while the smaller networks generalize the information.

These results indicate the 10 hidden node configuration to be best at identifying two jet events. For this reason, it was selected as the network used in all further research.

III. Network Pruning

A fundamental problem in using artificial neural networks is that there are more variables (weights) to be determined than there are parameters (training sets). It should be possible then to improve the determination of the variables, and thus network performance, if the number of variables are reduced. This process is referred to as pruning.

Pruning alters the neural network drastically, so care must be taken when determining which weights to eliminate. The obvious choice is to eliminate weights that are close to zero in magnitude, as they contribute little or nothing to the results. Assigning these weights a zero value eliminates them from being active in results calculation or learning. However, deciding which specific weights to eliminate is more complex.

For the 10 hidden layer node configuration used, only weights from the hidden layer to the output layer were considered for elimination. This was done to prevent the possibility of completely cutting off an input node which,

while it may not have been important for the input patterns the network was trained with, could be highly active for an unseen test event. However, there was a benefit to eliminating the weights to the output node; since there was only one output node, eliminating a weight effectively cut off the associated hidden layer node completely, and thus also the 960 weights that served as input to that node. No input layer nodes could be cut off with this method unless all the connections to the output were severed, which would render the network useless.

A pruning algorithm was then written to eliminate nodes to the output layer. Those weights which were smaller in magnitude than an adjustable percentage of the absolute average of the weights to the output layer were set to zero. The harshness of the prune was varied to see its effects.

The pruned network was then tested to show the effects of the pruning, and then, without altering the weights, retrained. The new network was then repruned and retrained in an iterative process to obtain the ideal network configuration.

The results of this pruning scheme were disappointing. The network showed a slow decline in performance as weights were eliminated and retraining occurred. As a result, the network was left unpruned for the remainder of modeling.

IV. Improving Data Simulation

In the original simulations, it was assumed that all the energy from a particle was deposited into a single calorimeter detector element. This is a poor assumption, as the energy is spread from the actual location in a nearly Gaussian fashion. This energy spreading is due to particle interactions inside the calorimeter detector element. For this reason, new data sets were developed using three different approximations for energy deposition to assess the ability of the network trained on data more closely resembling reality.

As a first attempt at energy smearing, it was assumed that each particle's energy deposition occurred in the center of the calorimeter detector element which it hit. Arbitrary percentages were chosen to distribute the energy such that the primary detector element was highly activated and surrounding elements were slightly activated.

A second method for energy smearing was to assume that all the particles impinged the calorimeter at the intersection of four calorimeter detector elements. This led to an equal distribution of the energy in each of the four elements around the intersection point.

These first two methods, while more spread than the original Monte Carlo data, do not provide a very accurate depiction of the real physical processes that are occurring. It is known that almost all of a particle's energy should be deposited within a relatively small circle around the impact point but at a certain depth into the calorimeter detector element. This depth was chosen as the position used for producing a more realistic simulation of data. The Gaussian energy distribution around the hit location was modeled by considering the distribution as a large number of individual hits with a corresponding small energy. Thus it was possible to distribute energy around the impact position by adding a random amount of energy to the detector element within the distribution radius. This was repeated until the total energy was distributed. These events were far more smeared.

V. Results of Data Smearing

A set of 1000 events was created from each of the spread approximations. These data sets were tested on the network trained with the rudimentary data. Table II summarizes the results. The network performed on- the same level for all smeared data as it did on the original data. This is indicates the smearing does not affect the pattern recognizing abilities of the network significantly. The Monte Carlo results show the network should perform well at identifying actual events.

TABLE II: Performance of Smeared Simulations

Center Smear Approximation			
	Number Events	Events Identified	Percentage
Min-Bias Events	500	472	95.8%
Two Jet Events	500	487	97.4%
Intersection Smear Approximation			
	Number Events	Events Identified	Percentage
Min-Bias Events	500	483	96.6%
Two Jet Events	500	480	96.0%
Monte Carlo Approximation			
	Number Events	Events Identified	Percentage
Min-Bias Events	500	489	97.8%
Two Jet Events	500	485	97.0%

VI. Potential Network Flaws

The disparity in energy deposition in the calorimeter between the minimum bias and two jet events, the two jet events are roughly a factor of 400 higher in energy deposition, brings into question the criterion for discrimination between the two classes of events. It is possible that the network was simply acting as a discriminator between the energies of the event types and not considering the patterns involved.

This possibility was investigated in two ways. First the events the network failed to identify were examined for any pathological behavior. No pathology was evident from the failures, particularly any relating to energy scale. A second more rigorous test was then conducted. The energies of the two jet events were scaled down by a factor of 100.

This reduced two jet energies to roughly the same order of magnitude as the minimum bias events. A set of 1000 events containing regular minimum bias and scaled two jet events was created for testing. If the network were only using energy levels to separate the events, results should show very poor results for scaled two jet data. This is not the case, as shown in Table III:

TABLE III: Results of Energy Scaling

	Number Events	Events Identified	Percentage
Min-Bias Events	500	482	96.4%
Two Jet Events	500	365	73.0%

The performance of the network trained on unscaled events is not up to the level of previous tests, but this is expected. The network is clearly using the energy as one criterion to distinguish events, but is obviously not ignoring the pattern of the events.

To demonstrate the network's ability to recognize patterns, it was retrained on the set of 1000 scaled events. An additional set of 250 scaled two jet and 250 minimum bias events was crated to test the network. Table IV details the results of these tests.

TABLE IV: Test of Pattern Recognition Ability

	Number Events	Events Identified	Percentage
Min-Bias Events	250	248	99.2%
Two Jet Events	250	236	94.4%

The results obtained are very close to those obtained by the original network. This indicates that the network's pattern recognition abilities pose no problem to the results.

VII. Conclusions

The results for two jet identification coupled with the evidence for pattern recognition indicate that a neural network of only 10 hidden layer nodes may be successfully used as a trigger. A hardware designed network should be capable of the speed required for the luminosity of SSC generation particle accelerators.

However, there are still areas that can be refined. Only one pruning algorithm was tried. A different pruning method may improve results. Event simulation could still be further improved.. In addition, the network could be trained using the most realistic data simulations, hopefully providing better results. As a final step, testing of the neural network built into a VLSI hardware trigger using real data should be done.

REFERENCES:

1. SDC Technical Design Report, SDC-92-201, Superconducting Super Collider Laboratory, Dallas Texas.

2. ISAJET F. E. Paige, S. D. Protopopescu, BrookHaven National Laboratory, Dec. 1988.

3. "Feasibility of Using Neural Networks as a Level 2 Calorimeter Trigger for Jet tagging," Proceedings of the International Conference on Computing in High Energy Physics '92," Annecy, France. September 21-25 1992, CERN 92-07, p650..
Eric Neis, University of Tennessee Science Alliance Summer Program Report, 1992.
Francis Starr, University of Tennessee Science Alliance Summer Program Report, 1993.

4. NeuralWare, Inc., Pittsburgh, PA 15276

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DATE
FILMED
5/4/94

END