Fermi National Accelerator Laboratory

FERMILAB-Pub-91/117

Neural Network Trigger Algorithms for Heavy Quark Event Selection in a Fixed Target High Energy Physics Experiment*

L. Gupta and A. M. Upadhye
Department of Electrical Engineering
Southern Illinois University
Carbondale, Illinois 62901

B. Denby
Fermi National Accelerator Laboratory
P.O. Box 500
Batavia, Illinois 60510

S. R. Amendolia and G. Grieco
University of Sassari and INFN Pisa, Italy

May 1991

* To be submitted to *Pattern Recognition*.

Operated by Universities Research Association Inc. under contract with the United States Department of Energy

NEURAL NETWORK TRIGGER ALGORITHMS FOR HEAVY QUARK EVENT SELECTION IN A FIXED TARGET HIGH ENERGY PHYSICS EXPERIMENT

Lalit Gupta and Anand M. Upadhye
Department of Electrical Engineering
Southern Illinois University
Carbondale, IL 62901

Bruce Denby
Fermi National Accelerator Laboratory
MS 234
Batavia, IL 60510

Salvator R. Amendolia and Giovanni Grieco
University of Sassari and INFN Pisa, Italy

ABSTRACT

The study of particles containing heavy quarks is currently a major topic in High Energy Physics. In this paper, neural net trigger algorithms are developed to distinguish heavy quark (signal) events from light quark (background) events in a fixed target experiment. The event tracks which are parametrized by the impact parameter D and the angle Φ of the track with respect to the beam line, vary in number and in position in the Φ-D plane. An invariant second order moment feature set and an invariant D-sequence representation are derived to characterize the signal and background event track patterns in the Φ-D plane. A 3-layer perceptron is trained to classify events as signal/background via the moments and D-sequences. A nearest neighbor classifier is also developed to serve as a benchmark for comparing the performance of the neural net triggers. Results indicate that the selected moment feature set and the D-sequence representation contain essential signal/background discriminatory information. The results also show that the neural network trigger algorithms are superior to the nearest neighbor trigger algorithms. A very high discrimination against background events and a very high efficiency for selecting signal events is obtained with the D-sequence neural net trigger algorithm.

SUMMARY

The study of particles containing heavy quarks is currently a major topic in High Energy Physics and, in this paper, neural net trigger algorithms are developed to distinguish heavy quark (signal) events from light quark (background) events in a fixed target experiment. The event tracks which are parameterized by the impact parameter D and the angle Φ of a track with respect to the beam line vary in number and in position in the Φ-D plane and cannot, therefore, be used as inputs to a neural network directly. This problem is overcome by deriving an invariant second order moment feature set and a D-sequence representation to characterize the signal and background tracks in the Φ-D plane. The moments feature set characterizes the dispersion of the tracks and the orientation of the tracks in the Φ-D plane. The D-sequence which is obtained through a simple set of transformations captures the track variations along the D-axis. A 3-layer perceptron is trained to classify events as signal/background via the normalized moments feature set and the D-sequences. The key to a successful study of heavy quark physics is a very high discrimination against background events and a high efficiency for selecting signal events. A training strategy is developed to keep the background misclassifications at a minimum. A nearest neighbor classifier is also developed to serve as a benchmark for comparing the performance of the neural net trigger algorithms. The very high efficiency obtained for rejecting background events and for selecting signal events clearly indicate that the selected moment feature set and the D-sequence representation contain essential signal/background discriminatory information. The results obtained also show that the neural net triggers are superior to the nearest neighbor triggers and the D-sequence neural net trigger is superior to the moments neural net trigger. It is important to note that the results obtained are very impressive as tests on randomly selected events indicate that, in many cases, it is impossible to visually distinguish between signal and background events from the track patterns in the Φ-D plane.

KEY WORDS: NEURAL NETWORKS, HEAVY QUARK PHYSICS, FIXED TARGET EXPERIMENT, TRIGGER ALGORITHMS

1. INTRODUCTION

This paper focuses on developing neural net trigger algorithms to distinguish a heavy quark event from a light quark event in a fixed target experiment. Currently, a major topic in High Energy Physics research is the study of particles containing the "heavy" quark flavors, "charm" and "beauty", and the search for the yet unseen top quark [1]. Particles containing heavy quarks, called charmed or beauty particles, or simply heavy flavor particles, can be produced in the interaction of a beam of high energy particles either with stationary nuclei (fixed target experiments) or with another beam of particles (collider experiments). In order to accurately catalog the properties of these particles, it is necessary to obtain relatively large samples of them, ranging from hundreds of events to millions or even hundreds of millions of events. At the same time, processes which produce heavy quarks (signal events) are very rare compared to less interesting background processes resulting in the "light" quarks (background events). The key to a successful study of charmed or beauty particles is thus a very high discrimination against background processes and a very high efficiency for selecting signal events [2].

The task is made even more difficult by the very high rate at which the data is produced. Since processes which produce heavy quarks are very rare, it is necessary to use the highest achievable beam intensities to ensure that the required data samples can be obtained in a reasonable amount of time (typically a year for an experimental run). Whatever the rate of signal production, however, the rate of background production will be much higher, almost a million times higher in the case of the NAxx fixed target test experiment at the

European Center for Nuclear Research (CERN) [2]. It is not feasible to record all of the events produced on storage media to evaluate offline. It is, therefore, necessary to make a decision online whether or not to record an event as it occurs. The amount of time available to make such a decision is typically in the order of ten microseconds.

A device called an experimental trigger which can determine, in a microsecond order time, whether a given event is likely to be a heavy flavor event or background event is required. Very fast experimental triggers have traditionally been built using custom made high speed digital and/or analog electronics coupled with high speed microprocessors. Such triggers, however, have in the past been used to make relatively "simple" trigger decisions, for example, to detect a charged particle with momentum above some threshold in a particular detector, or, to sum the total energy in an event and determine if this sum is above a threshold. Such simple criteria can hardly provide the necessary efficiency and background rejection for the high statistics study of heavy flavor events [1].

Recently it has been suggested that hardware neural networks could be used in experimental triggers for high energy physics experiments [3,4]. Feed forward neural networks are known to be able to perform a wide range of pattern recognition tasks, and the hardware implementations of such networks are very fast, with times to solution in the order of one microsecond. This paper describes a neural network approach to triggering based upon simulated data from the NAxx experimental project.

2. SECONDARY VERTEX CRITERION

The most promising criterion for distinguishing heavy flavor events from background events is the presence of a displaced vertex in heavy flavor events. High energy particle collisions typically produce many final state particles which emerge from a highly localized interaction point called the primary vertex. Final state particles containing heavy quarks are, however, unstable and will decay into daughter particles after travelling a small distance, producing a displaced, secondary vertex. There are background processes which can also produce secondary vertices such as strange particle production, but the distance travelled by heavy flavor particles before decay and the multiplicity of tracks produced in their decays are excellent criteria for distinguishing heavy flavor events from background events. In order to make use of the secondary vertex criterion, it is necessary to measure points along the trajectories of all primary and secondary tracks, to reconstruct the tracks and extract the parameters of the trajectories, and to reliably ascertain whether the tracks emerge from one or multiple vertices.

3. THE NAxx FIXED TARGET EXPERIMENT

In the NAxx project study, a proton beam p of energy 450 GeV/c^2 is extracted from the CERN super proton synchrotron and directed onto a thin metal target. Particles originating in the proton-nucleus interaction emerge in the forward direction at small angles to the beam direction and are detected in a set of silicon microstrip detectors as shown in Figure 1. The detectors are normal to the beam direction and measure the intercept of the track at each plane using metallized microstrips which collect the ionization produced by the particle in the silicon. Signals on the microstrip are amplified and pass via cables to a data acquisition system where they are digitized. The data acquisition system uses the strip information to calculate the hit positions, which in turn, are sent to a set of associative memory chips where the track parameters are found [5]. The strip information from each silicon plane passes serially through the associative memories, but the different planes are dealt with in parallel by the memories. The track parameters of detected tracks are available after all the hits have passed through the memories. The tracks may be parameterized by the impact parameter D which measures the distance of closest approach of the track to the origin, and Φ which is the angle of the track with respect to the beam line. The microstrips are oriented in three different directions to allow three dimensional reconstruction of the tracks, but here only one strip orientation is considered.

The Φ and D values for typical signal and background events are plotted in Figure 2. Φ is the abscissa and D is the ordinate of a track in the plots. The units of Φ and D are radians and centimeters respectively. In the background events, the tracks lie roughly on a horizontal line because all the tracks have come from a point in the thin target at $D=\hat{D}$ where $\hat{D}$ is the displacement of the beam particle from the origin. Often, there are tracks with large values of D in the signal events. This is because particles containing heavy quarks have travelled some distance from the interaction point and subsequently decayed into several more particles. These particles will have values of D that are different from those from the primary vertex. For each secondary vertex, in the limit of no measurement errors, the tracks from a secondary decay will lie upon a line whose slope with respect to the horizontal is proportional to the distance that the heavy flavor particle travelled from the origin (assuming the opening angle of the decay products is small, which is always the case).

What makes this a challenging pattern recognition problem is that in many cases, it is impossible to visually distinguish between a signal event and a background event from the track patterns in the Φ-D plane. Additionally, the number of tracks and the position of the tracks vary in the Φ-D plane. It is therefore clear that the classification must compensate for or be invariant to the number of tracks and the position of the tracks in Φ-D plane. The characteristics that prove to be useful for event classification are the variations in the impact parameter D and/or the dispersion of the tracks in the Φ-D plane.

4. NEURAL NETWORK APPROACH

Neural networks are very effective in solving pattern classification problems [6-12]. The neural network selected for the signal/background event classification problem is the multilayer perceptron. Generally, for a I-layered network which is fully inter-connected between adjacent layers, the output $X_i(k)$, $k=1,2,\ldots N_i$ of element k in layer i, $i=1,2,\ldots.I$ is given by

$$X_i(k) = f[\sum_{j=1}^{N_{(i-1)}} W_{(i-1)}(j,k)\, X_{(i-1)}(j) - \Theta_i(k)\,].$$

$X_0(k)$, $1 \le k \le N_0$ is the input vector; N_0 is the dimension of the input vector; N_i is the number of elements in the i^{th} layer; $W_0(j,k)$ is the connection matrix between the input and the first layer; $W_i(j,k)$, $1 \le i \le (I-1)$ is the connection matrix between the i^{th} and the $(i+1)^{th}$ layer; $\Theta_i(k)$ is the internal offset in the k^{th} element in layer i; f(.) is the sigmoid function.

The multilayer perceptron is trained under supervision using the backpropagation algorithm. Learning parameters of the backpropagation algorithm include a gain term ε and a momentum term α. The interconnection weights are adjusted recursively and the weights at the $(t+1)^{th}$ iteration are given by

$$W_{(i-1)}(j,k)\,|\,(t+1) = W_{(i-1)}(j,k)\,|\,t + \varepsilon\,\delta_i(k)X_i(j) + \alpha[\, W_{(i-1)}(j,k)\,|\,t - W_{(i-1)}(j,k)\,|\,(t-1)\,],\quad 1 \le i \le I$$

where

$$\delta_i(k) = X_i(k)[\,1 - X_i(k)][\,d(k) - X_i(k)\,],\quad i=I$$

$$= X_i(k)[\,1 - X_i(k)][\sum_{m=1}^{N_{(i+1)}} \delta_{(i+1)}(m)W_{(i+1)}(k,m)\,],\quad 1 \le i \le (I-1),$$

and $d(k)$ is the desired output of node $X_I(k)$.

The design of a multilayer perceptron for a particular classification problem involves determining:

(a) The number of layers.

(b) The number of elements in each layer.

(c) The format of the input to the network.

The complexity of the decision regions formed by the neural network can be increased by increasing the number of layers. When no prior knowledge of the decision region in the pattern space is assumed, a multilayer perceptron with 3 layers is a good choice because it is capable of forming arbitrarily complex decision regions. The number of elements in the input and output layers are governed by the input dimension and the number of classes respectively, and the number of hidden layer elements may be determined empirically. Generally, the output layer has one element per class. The input to the network must have a vector format and it is important to note that the dimension and ordering of the input vector must be fixed. A three layer perceptron is shown in Figure 3.

In the training stage, all interconnection weights are initialized to small random values with zero mean and the input training vectors belonging to the C different classes are presented cyclically until the net converges to the desired outputs. That is, the output $X_c = X_3(c)$ associated with class c, c=1,2,...,C is trained using the backpropagation algorithm to respond "high" when the input belongs to class c and respond "low" when the input does not belong to class c.

In the testing stage, the test vector is presented to the network input and the test event is assigned to the class c^* of the network output that yields the maximum value. That is

$$c^* = \arg\max_c [X_c], \quad c=1,2,\ldots,C.$$

A more complex rule based on thresholding the outputs to avoid ambiguous classifications may also be used. For correct classification, the thresholding scheme at the output layer requires a fixed combination of outputs to be above or below preset thresholds.

In developing neural net trigger algorithms for the signal/background classification problem, it is clear that the track data cannot be fed directly to the neural network because the number of tracks and the position of the tracks varies in the Φ-D plane. One possible approach to solve this problem is to extract a feature set from the track data which is invariant to the number of tracks and the position of tracks and use these features as inputs to the network. Another possibility is to derive a normalized representation of the track data which could be used as a neural net input. The derivation of an invariant feature set based on moments and a D-sequence transformation are described in the following sections.

5. MOMENTS

Moments and functions of moments have been widely used in pattern analysis and classification problems [13-17]. The $(p,q)^{th}$ moment of a function f(x,y) is defined as

$$m_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^p y^q f(x,y)dx\,dy, \quad p,q = 0,1,2,....,$$

and the $(p,q)^{th}$ central moment is defined as

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y)dx\,dy, \quad p,q = 0,1,2,....,$$

where, $\bar{x} = m_{10}/m_{00}$ and $\bar{y} = m_{01}/m_{00}$.

The order of a moment is defined as (p+q). Let F(Φ,D) represent the track pattern of an event in the (Φ,D) plane. To facilitate the computation of the moments, the tracks in F(Φ,D) are assumed to be "1" and all other points are assumed to be "0". The moments of F(Φ,D) are then estimated as

$$m_{pq} = \frac{1}{M}\Sigma\, \Phi^p D^q, \quad p,q = 0,1,2,....$$

$$\mu_{pq} = \frac{1}{M}\Sigma\, (\Phi-\bar{\Phi})^p (D-\bar{D})^q, \quad p,q = 0,1,2,.....,$$

where the summation is taken over all the M track points and (Φ,D) are the coordinates of a track point. Central moments are invariant to position in the (Φ,D) plane. Invariance to scaling corresponds to making μ_{00} equal to unity.

That is, the normalized central moments η_{pq} that are invariant to the track positions and invariant to the number of tracks are given by

$$\eta_{pq} = \mu_{pq}/(\mu_{00})^{\gamma}, \text{ where } \gamma = [(p+q)/2]+1 \text{ for } (p+q) = 2,3,....$$

In any classification technique based on moment features, it is important to select a set of moments which contains the most class-separation information as the set of all possible moments is not necessarily the best feature set for a given problem. From a computational point of view, it is advantageous to keep the moment feature set small. Generally, only the lower order moments are used in analysis and classification problems. Higher order moments tend to be unreliable due to their sensitivity to noise.

From the observations of the track patterns noted in Section 3, it is clear that the classification of an event into a signal or a background event can be accomplished by making use of moments which characterize track dispersion and track orientation information. Projection moments η_{02} and η_{20} characterize dispersion (variances) and η_{11} characterizes the orientation (covariance). Although higher order projection moments may contain additional information such as skew and kurtosis, they are not considered as features so as to keep the computation of the feature set as small as possible. After a further study of the second order normalized moments derived from the signal and background events, the moment η_{20} is dropped as a feature since it does not contain reliable class separation information. This is because the track angle Φ varies highly in both classes. The moment η_{02} is always positive whereas η_{11} may be positive or negative. In order to reduce the intraset variability in the feature set, only the magnitude of η_{11} is considered. Hence, the normalized second order moments $[(\eta_{02}), |\eta_{11}|]$ are selected as features to characterize the events.

The 3-layer perceptron moments classifier has two output elements X_s and X_b corresponding to the signal and background events and 2 input elements for the normalized moments $[(\eta_{02}), |\eta_{11}|]$. The hidden layer dimension was empirically chosen as 8. The network is tested with $[(\eta_{02}), |\eta_{11}|]$ of a test event in order to classify the event as a signal or background event. Since η_{02} and $|\eta_{11}|$ can take on extremely small values and present a problem in network training, the two normalized moments are scaled by a factor 10^6. A similar scaling operation is performed on the normalized moments of a test event.

6. D-SEQUENCES

Let the tracks in the signal and background events be represented by their coordinates in the Φ-D plane. That is, signal event i, i=1,2,...,... is represented by

$$(\Phi^s_{i,n}, D^s_{i,n}),\ n=1,2,\ldots,n_{i,s}$$

and background event j, j=1,2,...,... is represented by

$$(\Phi^b_{j,n}, D^b_{j,n}),\ n=1,2,\ldots,n_{j,b},$$

where, $n_{i,s}$ and $n_{j,b}$ are the number of tracks in the i^{th} and j^{th} signal and background events respectively. Let

$$s_i = \{s_i(1), s_i(2), \ldots, s_i(n_{i,s})\},$$

be a sequence which represents the distance between the tracks of signal event i and a horizontal line that passes through the centroid of the tracks. The k^{th} sample of s_i is given by

$$s_i(k) = | D^s_{i,k} - \bar{D}^s_i |,\ k=1,2,\ldots,n_{i,s},$$

where

$$\bar{D}^s_i = (m_{01})^s_i \,/\, (m_{00})^s_i,$$

and $(m_{pq})^s_i$ is the $(p,q)^{th}$ moment of signal event i.

Similarly, a sequence

$$b_j = \{b_j(1), b_j(2), \ldots, b_j(n_{j,b})\}$$

is derived to represent the track variations of the background event j along the D-axis. Although the sequences s_i and b_j contain the track variations in the D-axis, the variations do not follow any particular pattern along the Φ-axis. The degree of variability and hence the intraset distances in the sequences can be reduced through a rank ordering transformation. The sequences s_i and b_j are rank ordered such that

$$s_i(k) \le s_i(k+1),\ k=1,2,\ldots,(n_{i,s} - 1) \text{ and}$$

$$b_j(k) \le b_j(k+1),\ k=1,2,\ldots,(n_{j,b} - 1).$$

The rank ordered sequences can be used as feature vectors in the signal/background event classification problem. In order to use a 3-layer perceptron to classify signal and background events via the rank ordered sequences, the sequences need to have a fixed duration N (N is the dimension the input layer). Using a linear curve fit and a uniform resampling procedure [18,19], the sequences are normalized to have a duration N. The rank ordered and duration normalized sequences are referred to as "D-sequences". The i^{th} signal and j^{th} background D-sequences are denoted by

$S_i = \{S_i(1), S_i(2), \ldots, S_i(N)\}$ and

$B_j = \{B_j(1), B_j(2), \ldots, B_j(N)\}$

respectively.

The 3-layer perceptron D-sequence classifier has two output elements X_s and X_b corresponding to the signal and background events and 12 input elements (12 represents the mean number of tracks in the N signal and N background events selected to train the network). The hidden layer dimension was empirically chosen as 8. The network is tested with the D-sequence of a test event in order to classify the event as a signal or background event. In order to facilitate training, the D-sequences of the training events are scaled by a factor 100. A similar scaling operation is performed on the D-sequences of test events.

7. NEAREST NEIGHBOR CLASSIFIER

Given the moments feature vector and the D-sequence vector of the signal and background events, the nearest neighbor classification rule [20] may be used to classify the events. The performance of the nearest neighbor classifiers serve as benchmarks for evaluating the performance of the neural net classifiers. Let s_i, $i=1,2,\ldots,N_s$ be the i^{th} training signal event feature vector, b_j, $j=1,2,\ldots,N_b$ be the j^{th} training background event feature vector, t be the test feature vector, and d[.,.] be an appropriate distance measure. N_s and N_b are the number of signal and background events in the training set. Let $D_s^i = d[t,s_i]$; $D_b^j = d[t,b_j]$; $D_s = \min [D_s^i]$; $D_b = \min[D_b^j]$. The nearest neighbor rule assigns the test event represented by the vector t to the class of its nearest neighbor in the feature space. That is, the classifier assigns the test event to the class c^* given by

$$c^* = \arg\min [D_s, D_b], \quad \{s=\text{signal}, b=\text{background}\}.$$

The k-nearest neighbor classifier is an extension of the nearest neighbor classifier. The k-nearest neighbor rule assigns a test event to the majority of its k-nearest neighbors in the pattern space.

8. NETWORK TRAINING AND PERFORMANCE

A total of 2,500 signal and 2,500 background events were available for training and testing. In order to compare the performances, the training set was derived from exactly the same 100 randomly selected background and 100 randomly selected signal events for all trigger algorithms. Due to the similarity of many signal and background track patterns in the Φ-D plane, an overlap between the moments and the D-sequences belonging to the signal and background classes is expected. The overlap results in false alarms (background misclassifications) and misses (signal misclassifications). The performance of the trigger algorithms are measured by the false alarm and the miss probabilities. The number of misclassified background events divided by the number of background events tested gives an estimate of the false alarm probability P_f, and the number of misclassified signal events divided by the number of signal events tested gives an estimate of the miss probability P_m. A tradeoff between P_f and P_m can be approximately controlled by carefully selecting from the training set an appropriate training sub-set which controls the decision boundary formed by the neural network. As noted in Section 1, it is crucial to keep the false alarm rate as low as possible. The background misclassifications are kept at a minimum by selecting from the training set, a training sub-set which consists of the background points that outline the background boundary estimated from the training set and signal points that lie just outside this estimated background boundary on the signal side. The decision boundary formed by the neural network trained with the sub-set will thus be forced to approximate the estimated background boundary. Special attention is focused on excluding points that appear to be background outliers i.e. points with significantly different values which occur with extremely low probabilities. Additionally, signal points that fall on the background side of the estimated background boundary are excluded from the training sub-set.

The above training strategy is easily applied to determine the training set for the moments neural net trigger as the dimension of the feature space is 2. A total of 15 background points along the estimated background boundary, 15 signal points just outside this boundary, and 5 background and 5 signal points centered around the centroids of the background and signal clusters were initially used to train the network. One background point with unusually high moments was regarded as an outlier and was excluded from the training sub-set. The remaining 79 background points and 80 signal points in the training set were tested and the misclassified events were noted. All of the 5 misclassified background events were included in the training sub-set and the network was re-trained with the augmented training sub-set. To keep the number of signal and background training events equal, 5 signal points centered around the signal centroid were also added to the augmented training sub-set. The final training sub-set for the moments neural network, therefore, consisted of 25 background and 25 signal events.

Since the D-sequence feature space dimension is 12, the training set determination strategy is not directly applicable. Because the variations of the background events are typically smaller than that of the signal events, a rough grouping of background training events can be made by examining the rank 11 and rank 12 values of the events. In order to keep the background misclassifications at a minimum, 15 background D-sequences with high rank 11 and 12 values were selected to train the network for the background class. One background event with an unusually large rank 12 value was regarded as an outlier and was exluded from the training set. 15 signal D-sequences with rank 11 and 12 values slightly higher than the highest of the selected background D-sequences were chosen to train the network for the signal class. Additionally, 5 background D-sequences and 5 signal D-sequences with rank 11 and 12 values close to the mean rank 11 and 12 values were added to the training sub-set. The remaining 79 background events and the 80 signal events were tested and the misclassified events were noted. All of the 4 misclassified background events and 4 signal events with relatively large rank 11 and 12 values were included in the training sub-set and the network was retrained with the augmented training sub-set. The final training sub-set for the D-sequence neural network, therefore, consisted of 24 background and 24 signal events.

The test set consisted of the remaining 2,400 background and 2,400 signal events that were not used in determining the training sub-sets. The reference moments and D-sequence vectors for the nearest neighbor triggers were exactly the same vectors (final training sub-sets) used to train the moments and D-sequence neural networks. The false alarm and miss probabilities of all trigger algorithms are given in Table 1. In the k-nearest neighbor category, k was varied between 1 and 5. The best result was obtained for k=1 and this result is shown in the table. The Euclidean distance metric was used as a distance measure.

9. CONCLUSION

This paper describes neural net trigger algorithms designed to distinguish heavy quark events from a light quark events in a fixed target High Energy Physics experiment. Nearest neighbor triggers are also developed to serve as a benchmarks to evaluate the performance of the neural net trigger algorithms. The tracks produced by the events in the fixed target experiment vary in number and in position in the Φ-D plane and cannot, therefore, be used as inputs to a 3-layer perceptron directly. This problem is overcome by deriving an invariant moment feature set and a D-sequence representation to characterize the signal and background tracks in the Φ-D plane. The moment feature set characterizes the dispersion of the tracks along the D-axis and the orientation of the tracks in the Φ-D plane. The D-sequence which is obtained through a simple set of transformations captures the track variations along the D-axis. Results indicate that the moment feature set and the D-sequences contain essential signal/background discriminatory information. A training strategy is developed to keep the background misclassifications at a minimum. This training strategy is extremely useful in aiding the network to converge to a solution when pattern classes overlap. The results show that the neural net triggers are superior to the conventional nearest neighbor triggers. Highly acceptable false alarm and miss probabilities are obtained especially by the neural net D-sequence trigger algorithm. It is important to note that the results are very impressive as tests on randomly selected events indicate that, in many cases, it is impossible to visually distinguish between signal and background events from the track patterns in the Φ-D plane.

REFERENCES

1. A.J. Slaughter, N. Lockyer and M. Schmidt, "Proceedings of the Workshop on High Sensitivity Beauty Physics at Fermilab," Fermi National Accelerator Laboratory, Nov. 11-14, 1987.

2. S.R. Amendolia et al., "Research and development on fast trigger systems for heavy flavour triggering in fixed-target and collider experiments," Proposal to SPSC, CERN/SPSC 89-5, 13/1/1989.

3. B. Denby et al., "Neural networks for triggering," IEEE Transactions on Nuclear Science, vol. 37, No. 2, April 1990, 248-254.

4. S.R. Amendolia and B. Denby, "Ongoing Approaches to the triggering problem using neural networks," Proceedings of the Aachen Workshop, vol. 2, May, 1990.

5. L. Ristori and M. Dell'Orso, "VLSI for track finding," Proceedings of the International Conference on the Impact of Digital Microelectronics and Microprocessors on Particle Physics, Trieste, Italy, 28-30 March 1988, World Scientific, ISBN 9971-50-742-0.

6. D.E. Rumelhart and J.D. McClelland, "Parallel and distributed processing," vol. I, II, MIT Press, Cambridge, Mass., 1986.

7. R.P. Lippman, "An introduction to computing with neural nets," IEEE ASSP Magazine, 4-22, April 1987.

8. R.P. Gorman and T.J. Sejnowski, "Analysis of hidden units in layered network trained to classify sonar targets," Neural Networks, vol.1, 75-89, 1988.

9. L. Gupta, M.R. Sayeh, and R. Tammana, "A neural network approach to robust shape classification," Pattern Recognition vol. 23-6, 563-568, 1990.

10. L.J. Lang, A.H. Waibel and G.E. Hinton, "A time-delay neural network architecture for isolated word recognition," Neural Networks, vol.3, pp. 23-43, 1990.

11. L. Gupta, M.R. Sayeh and A.M. Upadhye, "Imposing a temporal structure in neural networks," Applications of Optical Engineering: Midwest 90, SPIE vol. 1396, 1991.

12. L. Gupta and A.M. Upadhye, "Non-linear alignment of neural net outputs for partial shape classification," Pattern Recognition (accepted for publication).

13. M.K. Hu, "Visual pattern recognition by moment invariants," IRE Trans. Info. Theory, 179-187, 1962.

14. S.A. Dudani et al., "Aircraft identification by moment invariants," IEEE Trans. Comput., vol. C-26, 39-45, 1977.

15. Y.S. Abu-Mostafa and D. Psaltis, "Recognition aspects of moment invariants," IEEE Trans. Pattern Anal. Machine Intell., vol. PAMI-6, 698-706, 1984.

16. L. Gupta and M.D. Srinath, "Contour sequence moments for the classification of closed planar shapes," Pattern Recognition, vol. 20-3, 267-272, 1987.

17. A. Khotanzad and Y.H. Hong, "Rotation invariant image recognition using features selected via a systematic method," Pattern Recognition, vol. 23-10, 1089-1101, 1990.

18. L. Gupta and M.D. Srinath, "Invariant planar shape recognition using dynamic alignment," Pattern Recognition, vol. 21-3, 1988.

19. L. Gupta and K. Malakapalli, "Robust partial shape classification using invariant breakpoints and dynamic alignment," Pattern Recognition, vol. 23-10, 1990.

20. R.O. Duda and P.E. Hart, "Pattern classification and scene analysis," New York: John Wiley & Sons, 1973.

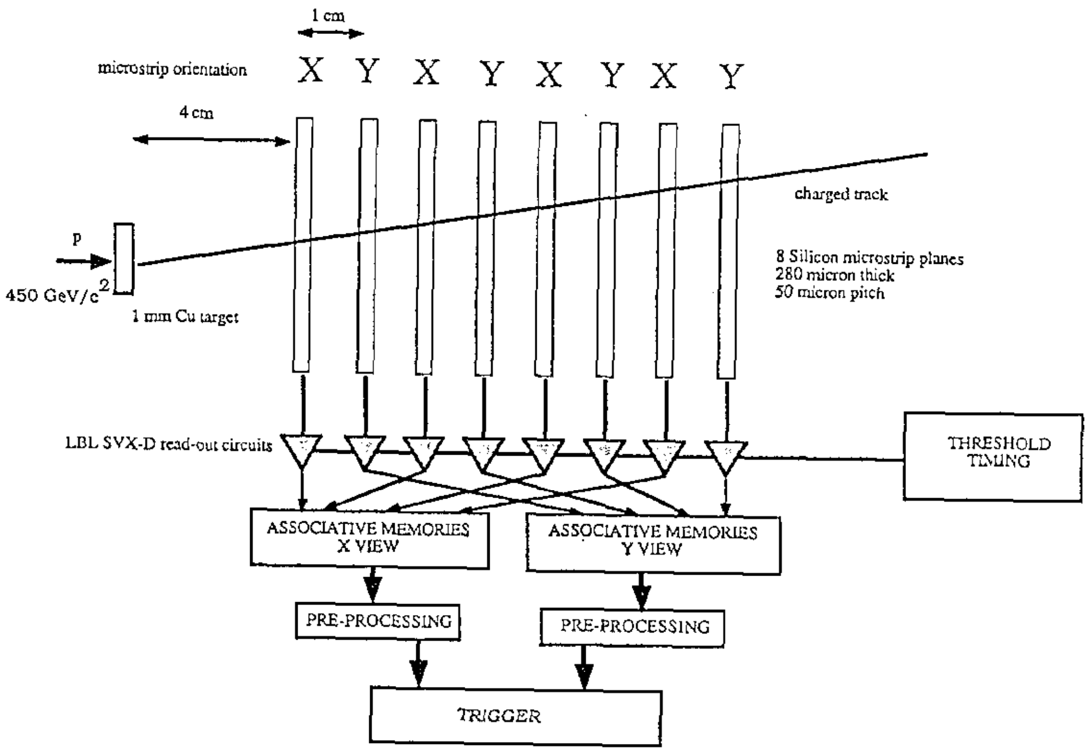

Fig. 1 Diagram of the NAxx test experiment.

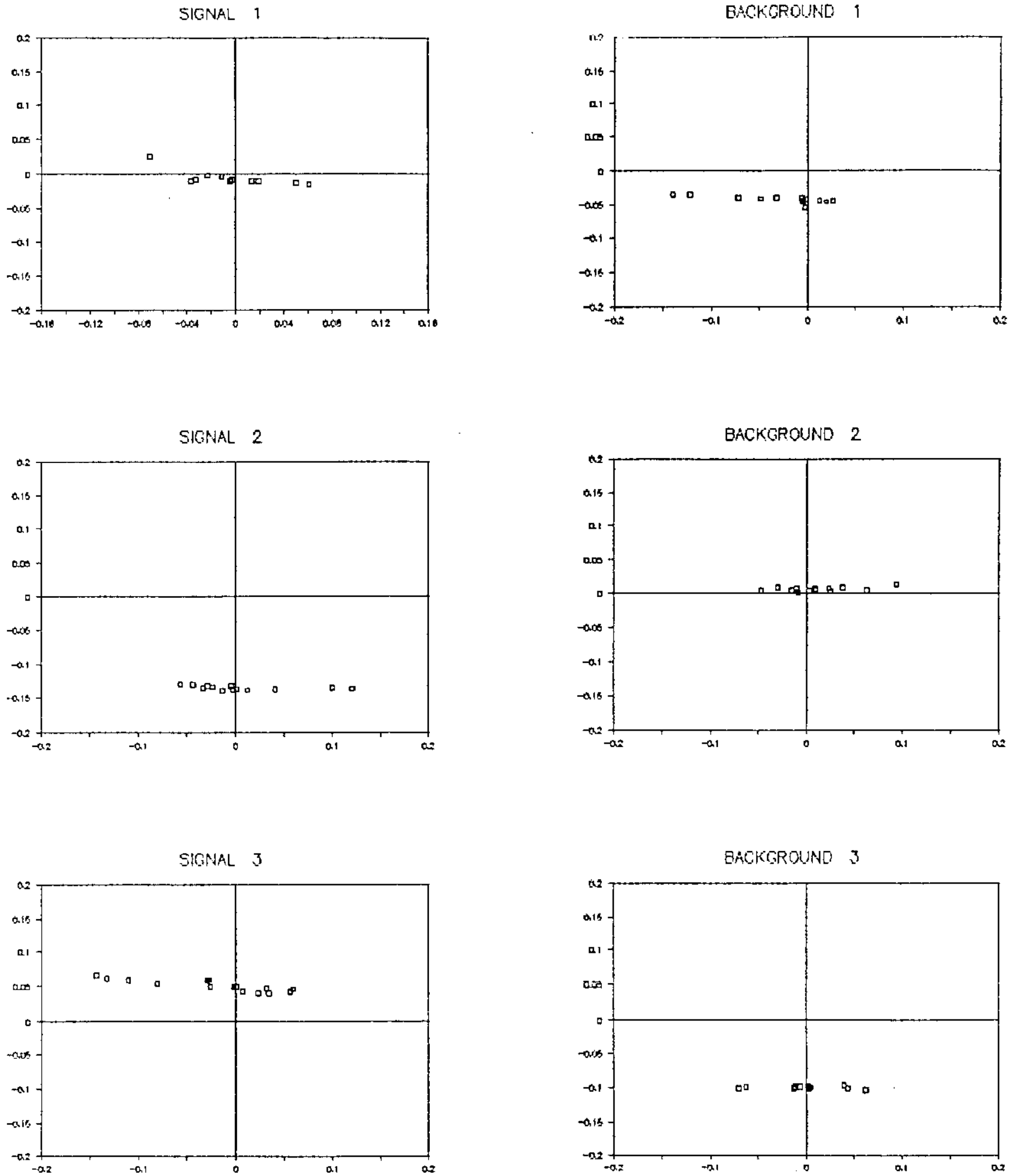

Fig. 2 Φ-D plots for typical signal and background events.

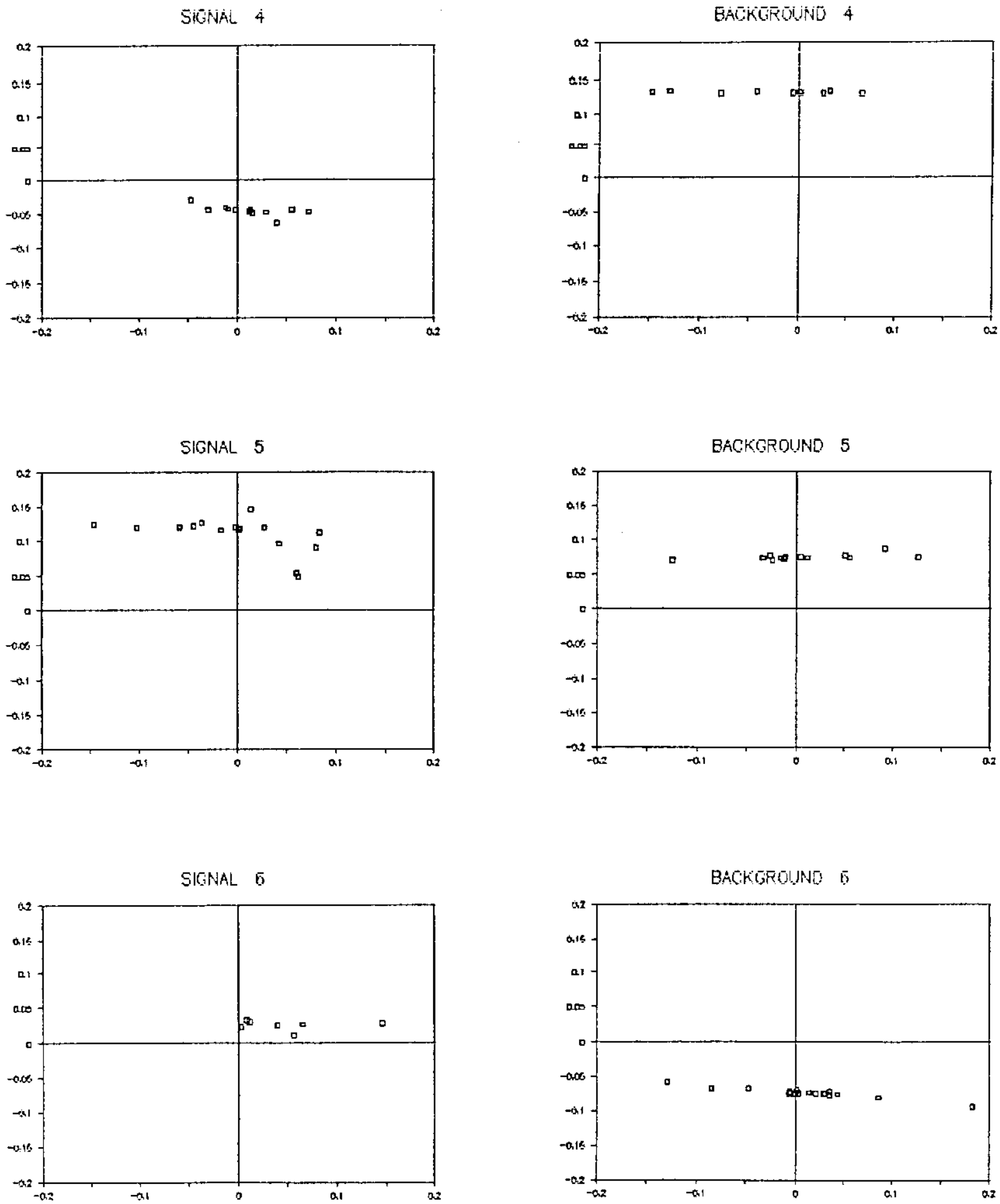

Fig. 2 (continued)

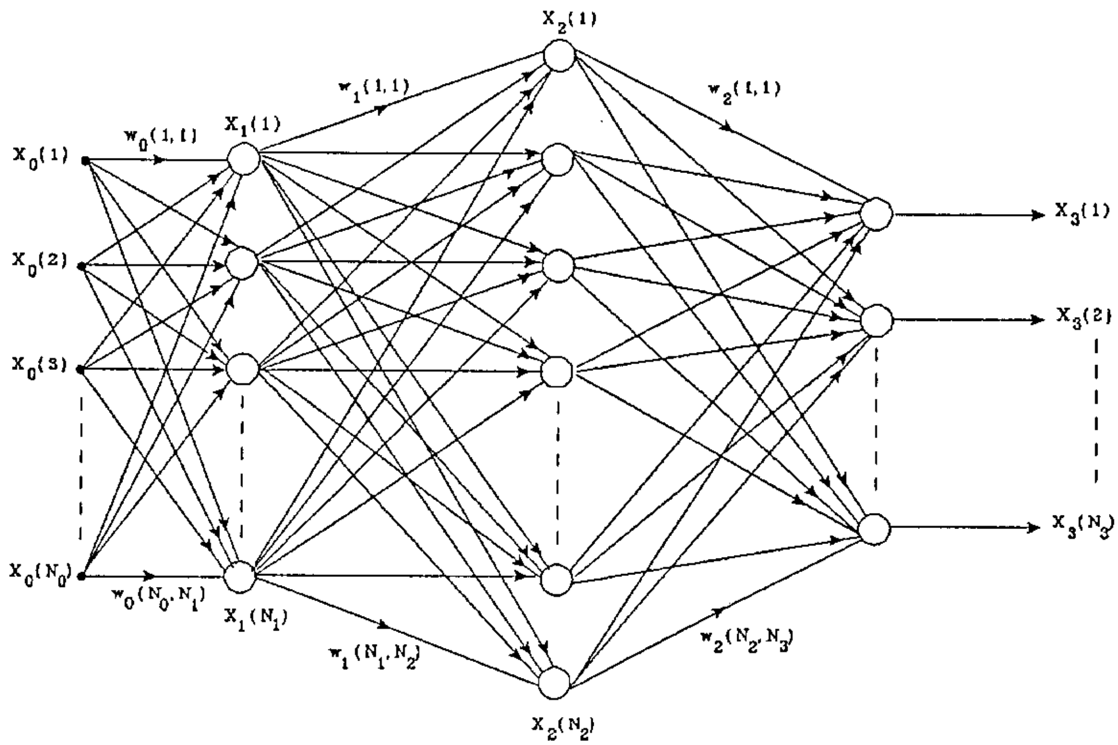

Fig. 3 Three-layer perceptron

		P_f	P_m
NEURAL NET	Moments	7/2400	461/2400
	D-sequences	1/2400	363/2400
NEAREST NEIGHBOR	Moments	27/2400	523/2400
	D-sequences	7/2400	382/2400

Table 1. False alarm and miss probabilities

LALIT GUPTA received the B.E. (Hons.) degree in electrical engineering from the Birla Institute of Technology and Science, Pilani, India (1976), the M.S. degree in digital systems from Brunel University, Middlesex, England (1981), and the Ph.D degree in electrical engineering from Southern Methodist University, Dallas, Texas (1986). Since 1986, he has been with the Department of Electrical Engineering, Southern Illinois University at Carbondale, where he is currently an Assistant Professor. His research interests include computer vision, pattern recognition, neural networks, and digital signal processing. Dr. Gupta serves as an associate editor for Pattern Recognition and is a member of the Pattern Recognition Society, the Institute of Electrical and Electronics Engineers, and the International Neural Network Society.

ANAND M. UPADHYE received the B.E., degree in electronics engineering from University of Bombay, India (1988). He is currently working on his M.S. degree in electrical engineering at Southern Ilinois University, Carbondale. His research interests include neural networks and computer vision.

BRUCE DENBY received the B.S. degree in physics from the California Institute of Technology (1975) and the M.S. and Ph.D degrees in experimental high energy physics from Rutgers (1977) and the University of California, Santa Barbara (1983) respectively. He has held positions at the European Center for Nuclear Research (C.E.R.N), Geneva, Switzerland, the Linear Accelerator Laboratory in Orsay, France, and, is currently at Fermi National Accelerator Laboratory, Batavia, Illinois. His research interests are experimental high energy physics at hadron colliders and applications of neural networks to triggering and data analysis for high energy physics experiments.

SALVATOR R. AMENDOLIA is an Associate Professor of Physics at the University of Sassari, Italy and a senior researcher with I.N.F.N. of Italy. Since 1971, he has been working at CERN in Geneva, on the total proton-proton cross-section experiment at ISR, on the NA1 charmed-mesons photoproduction experiment at SPS (where he has been the manager of the DAQ and Trigger System), and on the Aleph experiment at LEP (as head of the Online design group, beyond working on data analysis). He is presently involved in the design of the SDC experiment at the Superconducting Super Collider in U.S.A. He is spokesman for NAxx, an R&D project funded by I.N.F.N. Italy, which studies novel trigger techniques for High Energy Physics experiments with the use of Neural

Networks and advanced VLSI technologies. He is the inventor and scientific co-spokesman of a proposed 4 million dollar EEC project on the automatic recognition of NMR spectra with the use of neural networks.

GIOVANNI GRIECO is a graduate student associated with I.N.F.N. at Pisa. He has worked on his thesis with Professor S.R. Amendolia on the NAxx R&D project on the development of new trigger techniques for High Energy Physics. He recently joined the CDF collaboration at FNAL, where he plans to work on the Silicon Vertex Detector.