

SELECTION OF PHOTON GLUON FUSION EVENTS IN DIS

KATARZYNA KOWALIK, EWA RONDIO

The A. Soltan Institute for Nuclear Studies[†]
Hoża 69, 00-681 Warsaw, Poland

ROBERT SULEJ, AND KRZYSZTOF ZAREMBA

Warsaw University of Technology[‡]
Koszykowa 75, 00-662 Warsaw, Poland

(Received July 24, 2001; revised version received September 25, 2001)

A selection of the Photon Gluon Fusion (PGF) process with light quarks for deep inelastic scattering events is presented. This process is directly sensitive to gluon polarization and our goal is to find out the most effective selection on a sample of events simulated for the SMC experiment. We compare two general multi-class classification methods — Bayes method and neural network with a conventional selection procedure. The neural network algorithm presented here is a modification of method belonging to the family of directional minimization algorithms. This method is convenient and effective for photon gluon fusion selection and determination of gluon polarization. Finally we present the estimation for precision of gluon polarization for neural network method.

PACS numbers: 13.10.+q, 13.85.Fb, 07.05.Mh

1. Introduction

The spin structure of the nucleon has been studied in polarized Deep Inelastic lepton–nucleon Scattering (DIS) for quite a long time. The experimental observation by EMC [1], that a surprisingly small fraction of the nucleon spin is carried by quarks has had major influence on the more recent

[†] This work was supported in part by the Polish State Committee for Scientific Research (KBN) SPUB Nr621/E-78/SPUB-M/CERN/P-03/DZ298/2000 and grant Nr 2 P03B 113 19.

[‡] Supported by the Polish State Committee for Scientific Research (KBN) SPUB Nr134/E-365/SPUB-M/CERN/P-03/DZ299/2000.

spin physics. Many experiments were performed to confirm this result [2–7] and check it on different targets. More measurements are in progress and in preparation [8–10]. Several theoretical ideas were proposed [11] to explain this observation. They were based either on large negative sea quark polarization or on contribution from polarized gluon and orbital angular momentum of quarks and gluons.

To check which of them is responsible for making up the nucleon spin one should determine a fraction of nucleon spin carried by gluons. Information about this quantity can be obtained indirectly from the dependence of the structure function g_1 on four-momentum transfer squared, Q^2 . In this analysis parton distributions (quarks (q) and gluons (G)) are fitted using QCD evolution to the measurements of the spin dependent structure function g_1 [12,13]. However, due to the large number of theoretical assumptions in this method, a direct measurement of gluon polarization seems to be the best way of verifying the spin structure of the nucleon.

The first measurement of the polarized gluon density was performed by E704 experiment at Fermilab but unfortunately the results did not allow to distinguish between different theoretical models. The Hermes experiment at DESY also has measured gluon polarization. In their case the precision is severely limited and the interpretation is not clear due to low momentum transfer. The experiments STAR/PHENIX at RHIC, COMPASS at CERN and E161 at SLAC are planned to take data next years. Their precision on $\delta(\Delta G/G)$ will be better than indirect measurements. Also the possibility of using colliding polarized proton and electron beam at HERA has been discussed.

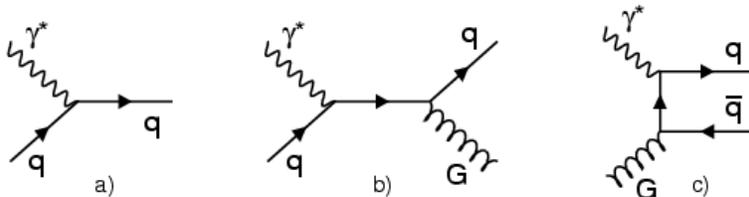


Fig. 1. Lowest order diagrams for DIS γ^*N scattering: a) virtual photo-absorption (LO), b) gluon radiation (Compton diagram), c) photon gluon fusion (PGF).

In DIS the leading order process, the virtual photo-absorption, does not allow direct access to the gluon distribution, since the virtual photon does not couple directly to the gluon. Hence the observation of higher order processes is an alternative solution to get gluon distribution compared to the indirect method based on QCD evolution.

Such a direct measurement is possible via the Photon Gluon Fusion (PGF) process. The Feynman diagram of this process is presented in Fig. 1(c)

together with leading order (Fig. 1(a)) and gluon radiation — Compton diagrams (Fig. 1(b)). Both PGF and Compton processes are of the same order in strong coupling constant, α_S , so they have lower contribution to DIS cross section than leading order diagram.

Since the frequency of PGF is small in comparison to all processes, finding a signature to tag this process is very important. The most straightforward way of searching for PGF is charm production signalled either by detection of charmed particles (especially D^0) or by production of J/ψ . For such processes the contribution from leading order diagram is small because contents of the charm quark in the nucleon is practically negligible. Due to the large mass of the charm quark the contribution via the fragmentation processes is also low. For the same reason charm pair production in PGF is suppressed. Therefore it is important to find a method which allows to identify PGF in case of light quarks production. Tagging the PGF process by observation of hadrons with large transverse momentum (p_T) in the final state gives such possibility. The p_T is calculated with respect to the virtual photon direction.

In the leading order process the contributions to hadrons p_T are the intrinsic k_T of quarks in the nucleon [14] and the fragmentation process. Thus most hadrons have small transverse momentum. The opposite situation occurs for Compton and PGF where hadrons mainly acquire transverse momentum from primary produced partons. For this reason the requirement of observation of two hadrons with large transverse momentum can enhance the contribution of the PGF process in the selected sample. This idea has been recently discussed in [15]. It was applied to determine the gluon polarization from photoproduction data in the Hermes experiment [16].

In this paper we present different approaches to the selection of PGF with light quark production. In Section 2 a short theoretical description of the determination of gluon polarization is presented. Methods based on Bayes classification and the neural network are discussed in Section 3. The Monte Carlo sample used in the tests is described in Section 4. In Section 5 we compare results obtained in different methods:

1. Bayes classification,
2. the neural network,
3. traditional cuts on kinematic variables.

For the best classification method we determine the conditions providing the optimal precision on the gluon polarization determination.

2. Formalism

The experimentally measured quantity is the spin cross section asymmetry defined as the ratio of polarized and unpolarized cross sections:

$$A^{lN} = \frac{\Delta\sigma}{2\sigma} = \frac{\sigma^{\uparrow\downarrow} - \sigma^{\uparrow\uparrow}}{\sigma^{\uparrow\downarrow} + \sigma^{\uparrow\uparrow}}, \quad (2.1)$$

where $\sigma^{\uparrow\downarrow}$ and $\sigma^{\uparrow\uparrow}$ are the cross sections with antiparallel and parallel orientation of beam and target polarizations, respectively. The measured A^{lN} asymmetry is related to the polarized distribution of quarks $\Delta q = (q^\uparrow - q^\downarrow)$ and gluons $\Delta G = (G^\uparrow - G^\downarrow)$. Here, the arrows correspond to antiparallel and parallel configuration of parton and nucleon spin.

The unpolarized cross section can be expressed as:

$$\sigma = F \otimes \hat{\sigma} \otimes D \quad (2.2)$$

and the polarized one as:

$$\Delta\sigma = \Delta F \otimes \Delta\hat{\sigma} \otimes D, \quad (2.3)$$

where F and ΔF are unpolarized and polarized quark or gluon distribution functions. The $\hat{\sigma}$ and $\Delta\hat{\sigma}$ symbols are, respectively, the spin-independent and spin-dependent partonic hard-scattering cross sections. The function D describes the fragmentation of partons into hadrons and the symbol \otimes stands for convolution.

The full expression to calculate unpolarized (σ) and polarized ($\Delta\sigma$) cross sections of inclusive production of two hadrons with large p_T in the final state consists of three terms since all processes shown in Fig. 1 should be taken into account.

We can extract the gluon polarization $\frac{\Delta G}{G}$ by inserting the full expression for σ and $\Delta\sigma$ into Eq. (2.1). The final expression reads:

$$\frac{\Delta G}{G} \langle \hat{a}_{LL} \rangle^{\gamma^* G \rightarrow q\bar{q}} R_{\text{PGF}} = A_1 \langle \hat{a}_{LL} \rangle^{\gamma^* q \rightarrow q} R_{\text{LO}} + A_1 \langle \hat{a}_{LL} \rangle^{\gamma^* q \rightarrow qG} R_{\text{Compton}} - A^{lN \rightarrow hhX}, \quad (2.4)$$

where $\langle \hat{a}_{LL} \rangle = \frac{\Delta\hat{\sigma}}{\hat{\sigma}}$ is the partonic asymmetry for the hard scattering processes, the ratios R refer to the contribution of each process shown in Fig. 1 to the total cross section, $A^{lN \rightarrow hhX}$ is the measured asymmetry for the selected events and A_1 is the virtual photo-absorption asymmetry which is well known from inclusive experiments.

The values of $\langle \hat{a}_{LL} \rangle$ are calculated for each event using the matrix element for partonic processes in the program POLDIS [17]. Their behavior as the

function of scattering angle in the parton–photon c.m. significantly depends on the type of process. For the Compton process the asymmetry $\langle \hat{a}_{LL} \rangle^{\gamma^* q \rightarrow q G}$ is positive in contrast to the one for PGF — $\langle a_{LL} \rangle^{\gamma^* \bar{G} \rightarrow q \bar{q}}$ which is negative and almost two times bigger in the relevant kinematic region.

As can be seen from Eq. (2.4) the precision of the gluon polarization determination depends on the statistical precision of the measured asymmetry $A^{IN \rightarrow hhX}$ and on the contribution of background processes (R_{LO} and $R_{Compton}$) to the final sample. Since the background contribution introduces the major systematic uncertainty in the evaluation of $\frac{\Delta G}{G}$, the goal of the selection is to obtain a sample with maximal number of PGF events and minimal contributions of background processes. The average value of $\langle \hat{a}_{LL} \rangle$ and the ratios R for the final sample are estimated from Monte Carlo simulations.

The criteria to judge the selection are based on two numbers: purity, which is the fraction of wanted PGF events in the finally selected sample, and efficiency defined as the fraction of PGF events from input which survives the selection.

3. Classification techniques

Conventional approaches, based on cuts on event and hadron variables, have two serious weak points. The first one is the laborious search for the optimal set of cuts, often done with a trial and error technique. The second disadvantage is treating parameters independently, even when obvious correlations do occur. Also, there is no possibility of continuous balance between selection efficiency and sample purity when using cuts.

We show two alternative techniques of selecting PGF events: neural network and, for comparison, a pure statistical technique based on the Bayes theorem, which is known as a “close to optimum” standard in multi-class classification problems.

3.1. Bayesian approach

This technique consists in computing the conditional probability, with which a given event belongs to the distribution of the process of interest (all other processes are treated as background). This probability is given by the expression:

$$P(G_n | \mathbf{X} = (x_1, \dots, x_k)) = \frac{g_n(x_1, \dots, x_k) p_n}{\sum_{i=1}^m g_i(x_1, \dots, x_k) p_i}, \quad (3.1)$$

where: G_n — is the process of interest; \mathbf{X} — the event feature vector; $g_i(\mathbf{X})$ — the i -th process distribution; p_i — the *a priori* probability of i -th process; m — the total number of processes.

The bayesian classification described in this section can be applied for the PGF process selection. An event is classified as PGF if the obtained probability exceeds a fixed threshold Tr : $P(\text{PGF}|\mathbf{X}) > \text{Tr}$. Changing this threshold allows us to increase the selection purity at the cost of classified set's statistics, or, on the other hand, we can get higher selection efficiency with lower purity of classified sample.

This technique would give an optimal solution on one condition: it is necessary to know the real probability distributions and the 'a priori' probabilities for all processes in the experiment. This is not true in our case and we can only estimate their values from a limited set of events (called later "a training set"); see [18] for more details on this technique.

A good estimator of $g(\mathbf{X})p$ for a given process (convergent to the real value with growing number of events in the training set) is the sum of functions attached to each event corresponding to this process in the training set. These functions (called also "potential functions", [19]) should have a maximum at the point with coordinates equal to the feature vector of a given event and their values should decrease with growing distance from that point, so that the event has strong influence on the estimator value only in its close surrounding. Gaussoid functions comply with these requirements quite well. The estimator defined in this way takes the form:

$$\hat{g}_i(\mathbf{X})\hat{p}_i = \sum_{j=1}^{n_i} \exp\left(-\frac{1}{r}d(\mathbf{X}, \mathbf{Y}_j)^2\right), \quad (3.2)$$

where n_i is the number of events of the i -th process in the training set; r — is the Gaussian function width; $d(\mathbf{X}, \mathbf{Y}_j)$ — indicates the distance between the feature vector \mathbf{X} and the feature vector \mathbf{Y}_j of the j -th event from the training set (corresponding to the i -th process): $d = \sqrt{\mathbf{X}^T \mathbf{Y}_j}$. In Eq. (3.1) we calculate ratios of $g(\mathbf{X}) \cdot p$ values so there is no need to normalize the estimator values.

The Gaussoid function width (parameter r) determines the radius of influence of a single event on the estimator shape. It is optimized with an algorithm that minimizes the Mean Squared Error (MSE). This error is calculated on a testing set of events as the mean value of the squared deviation between correct answer and Bayes probabilities obtained with Eq. (3.1):

$$\text{MSE} = \frac{1}{n} \sum_{j=1}^n [d_j - P(\text{PGF}|\mathbf{X}_j)]^2, \quad (3.3)$$

where d_j is a correct answer for event \mathbf{X}_j from a testing set of n events ($d_j = 1$ for PGF process, $d_j = 0$ for background process). In practice,

a good algorithm (quickly convergent to optimum value of r) is the one that assumes a parabolic dependence $\text{MSE}(r)$ in minimum surroundings.

Because of the very long calculation time in the selection of the PGF process, it was necessary to make a simplification: the Gaussoid function width r was assumed to be the same in all directions of a feature space. According to this simplification, data in the training set were normalized to get similar numerical values of all parameters *i.e.* to prevent features with large standard deviation from dominating the r optimization process.

A serious and important problem in bayesian classification is the complexity of calculations. Distribution's estimator values have to be computed each time we classify an event from the testing set, during both optimization and classification processes. It means, according to Eq. (3.2), that about (Nm) multiplication and accumulation operations and N calculations of exponential function should be done to classify one event for a training set consisting of N events described with m parameters. The value of m gives the number of components of the feature vector and depends on the number of variables used to describe an event (in our case $m = 3$). The number of events in the training set (N) is chosen as a compromise between the quality of estimator (statistics of processes) and the calculation time. The required value of N dramatically increases when new variables appear in a feature vector, which makes this technique very time-consuming.

3.2. Neural network

The neural network application is an alternative method of event classification. A simple network, in feed-forward configuration (Fig. 2), was used in the project.

Neural networks consist of multiple, simple processing units (artificial neurons) interconnected by large number of weighted connections. In the feed-forward configuration neurons are arranged in a few layers and data flow is strictly feed-forward *i.e.* from the first (input) layer, through the number of hidden layers to the output.

Each neuron receives an input signal which is a sum of signals connected to its inputs multiplied by connection weights and generates an output signal which is a certain function of the input. A so-called sigmoid function is used in most of cases. Eq. (3.4) describes the neuron operations shown in Fig. 3.

$$\text{out} = \frac{1}{[1 + \exp[-w_0 - \sum_{i=1}^n x_i w_i]]}. \quad (3.4)$$

A neural network is trained by feeding a set of teaching patterns and changing the w_i weights according to some learning rule. In the supervised learning, which was applied in present project, training vectors are supplemented

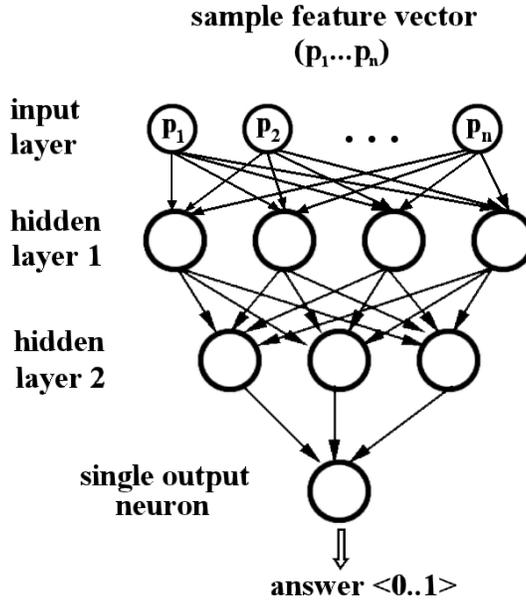


Fig. 2. Network topology.

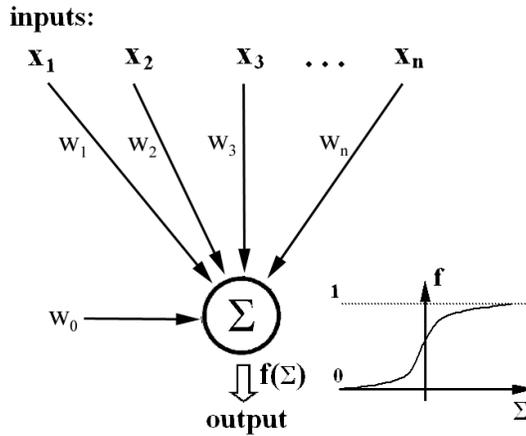


Fig. 3. Neural unit.

by matching output patterns. The weights are changed usually with a gradient descent method which iteratively minimizes an error function which is defined as the mean square error between desired and actual outputs of all neurons.

The structure (topology) of the network depends on the application (see Refs [18, 20, 21]). In the classification task presented in this paper, a net with two hidden layers was used. The number of neurons in the input layer

corresponds to the number of parameters used in the data processing. The output layer contains a single neuron. The response of this neuron varies between 1 (the desired output of PGF) and 0 (other processes). The answer of the net may be interpreted as a conditional probability. The threshold, which we can apply to the neural network output, corresponds to the Tr threshold for the bayesian classification.

As mentioned above, a training set of patterns is needed for teaching the network. Each pattern corresponds to the feature vector of an event. There should be a couple of times more patterns (of each process) in the training set than weights in the network. The most commonly used gradient descent training algorithm is the BP method (error Back Propagation). However, there are more efficient techniques such as conjugated gradients. The training algorithm used in this project is a modification of this technique. The applied algorithm can be decomposed into the following steps:

1. Initialize weight values as random numbers (there is no *a priori* knowledge where the starting point should be): point "0" in Fig. 4.
2. Calculate the negative gradient: direction of the greatest inclination of the MSE surface (mean squared deviation between answers obtained from the net and correct answers for all events from the training set).
3. Move through the weights space in this direction (with relatively big initial steps) until the angle between current negative gradient and searching direction exceeds 90° : points "1" and "2".
4. Reverse searching direction and shorten the step to half of its previous value.

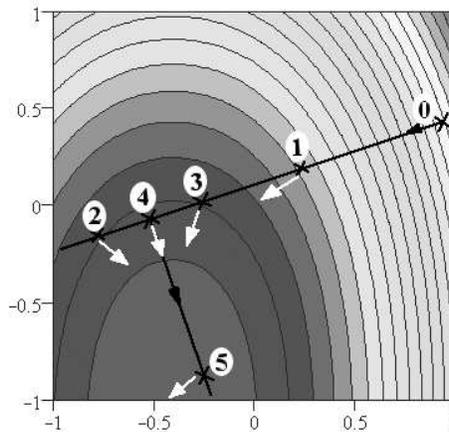


Fig. 4. Illustration of the training algorithm.

5. Repeat (3) and (4) until minimum step value is reached: points “3” and “4”.
6. Change searching direction to the current negative gradient direction (nearly perpendicular to the previous one): point “5”.
7. Repeat (3)–(6) until the terminal condition (MSE value, number of training set presentations) is satisfied.

This algorithm is similar to the well-known conjugated gradient algorithm (see Refs [18, 22–24]), which belongs to the family of directional minimization algorithms. In this family, basically, the point of minimum error is searched for in a fixed direction. Then the searching direction is changed according to some rules. In the conjugated gradient algorithm, the new direction is calculated as:

$$\mathbf{d}_n = \mathbf{g} + \gamma \cdot \mathbf{d}_{n-1}, \quad \gamma = \frac{(\mathbf{g} - \mathbf{d}_{n-1})^T \cdot \mathbf{g}}{\mathbf{d}_{n-1}^T \cdot \mathbf{d}_{n-1}}, \quad (3.5)$$

where \mathbf{g} is a current negative gradient and \mathbf{d}_{n-1} is previous searching direction. Eq. (3.5) assumes a squared form of error surface in minimum surrounding and, if the MSE surface complies with this condition, the algorithm assures the smallest number of direction changes.

The conjugated gradient algorithm was also tested and serious disadvantages were observed. This algorithm is very sensitive to the value of the initial step (used in searching for the minimum in fixed direction). A too big value makes the algorithm unstable. It forces starting from tiny steps each time the algorithm changes searching direction. This causes that, in spite of lower number of direction changes, we have to calculate often a comparable number of iterations to reach the same MSE level as with the first presented algorithm (Fig. 5(a) and Fig. 5(b)). Difficulties in finding the optimal initial step value were the reason of giving up further tests with the conjugated gradient algorithm. Our modified algorithm accepts a wide range of initial step values. Bigger values speed up the training process (Fig. 6) and there is no danger of making the process unstable (a very large initial step only slows down the training process).

Because of the possibility of finding a local minimum (the MSE surface may have a lot of them) it is good to train the network with different initial weight values and, if the network gives similar results in these trials, it is a fine proof of finding a global minimum. The presented algorithm is capable of escaping from local minima in many cases and there was no case of significant differences in results obtained by networks trained with randomized initial weights. The training process should be stopped when no significant change in the MSE value is observed.

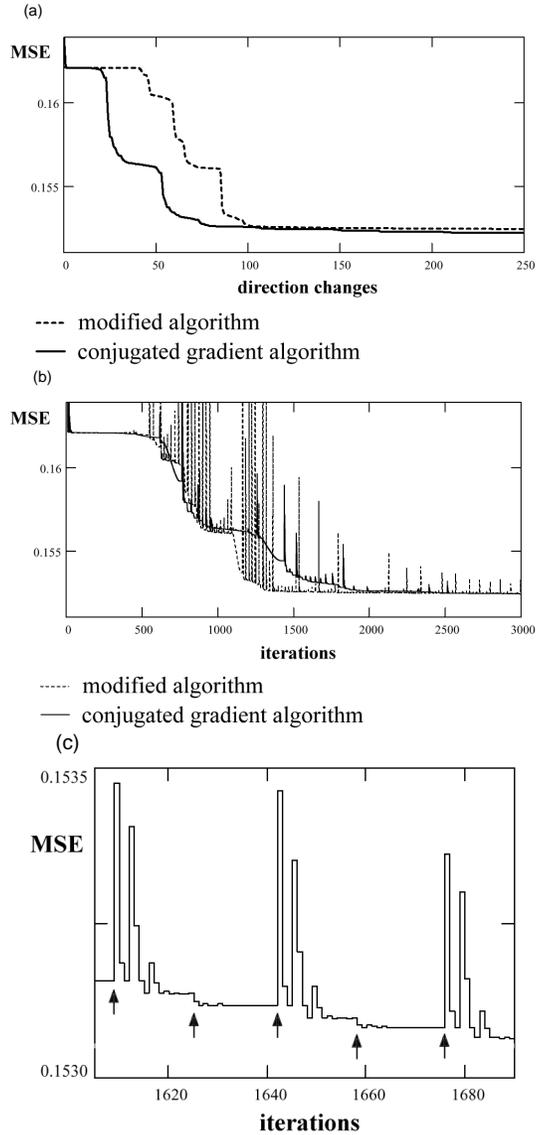


Fig. 5. Training process: (a) MSE as a function of direction changes (MSE values before direction changing are plotted only); (b) MSE as a function of training iterations; (c) enlarged fragment of the MSE diagram (arrows mark the iterations when the search direction is changed). An initial step of 0.01 (maximum stable value) was used for the conjugated gradient algorithm. An initial step of 0.04 was used for the modified algorithm.

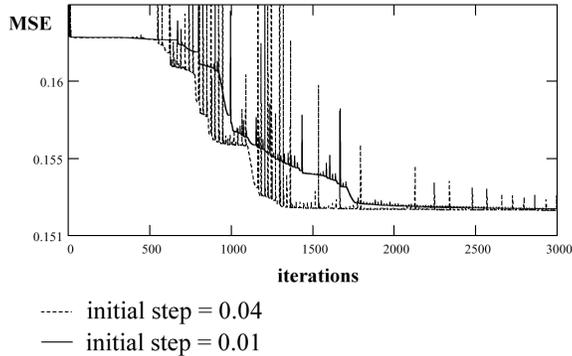


Fig. 6. Training process speed for modified algorithm for two different initial step values.

One of main advantages of the network is its computation simplicity. The number of required calculations depends on the network structure, which is given by:

- H_1, H_2 , the number of neurons in hidden layers; sufficient values for PGF events classification are $H_1 = 5 \div 6$ and $H_2 = 3$,
- m , the size of the input event feature vector, $m = 3 \div 8$.

For the classification of a single event the network requires

$$H_1(m + 1) + H_2(H_1 + 1) + H_2 + 1$$

accumulation and multiplication operations and $H_1 + H_2 + 1$ calculations of the neuron output function, which is usually implemented as segment linear approximation. It means that in practical applications classification of events with the neural network is about 10^3 faster than with the bayesian technique. Also, the simplicity of classifying of events makes the training time much shorter than the optimization of estimation parameters (Gaussian function width) for the bayesian technique.

4. Sample description

The studies were performed for the specific setup of the SMC experiment realized in the high energy muon beam at CERN. The polarized muons of 190 GeV scatter off polarized nucleons producing hadrons in the final state. The interaction can be described by following kinematic variables: ν is energy transfer to the virtual photon, $y = \nu/E_b$ where E_b is beam energy, Q^2 is the squared mass of the virtual photon and $x_{Bj} = Q^2/2M\nu$ where M is the proton mass. Using the above variables the invariant mass of the final hadronic state W^2 can be defined.

The scattered muon and produced charge particle were measured in SMC open magnetic spectrometer. The minimal energy of hadron which can be detected in spectrometer is around 5 GeV.

The experimental setup was simulated using the LEPTO [25] generator with parameterizations of parton distributions at leading order taken from [26]. The spin dependent effects were calculated in POLDIS [17] with a consistent polarized parton distribution set [27]. In generation the following kinematics cuts were applied:

$$0.003 < x_{Bj} < 0.7,$$

$$0.05 < y < 0.93,$$

$$1 \text{ GeV} < Q^2 < 100 \text{ GeV},$$

$$6.5 \text{ GeV}^2 < W^2 < 1000 \text{ GeV}^2,$$

$$10 \text{ GeV} < \nu < 190 \text{ GeV}.$$

The generated sample is composed of 88% leading order process, 8% Compton and 4% PGF process. For each event the muon track was followed through the magnet and trigger conditions were checked. About 33% events fulfill the trigger conditions. The fraction of events which have at least two hadrons in the acceptance is 52%. Finally only events with at least two hadrons with p_T above 0.7 GeV were kept. With this selection we are left with 4% of generated event sample. The fraction of PGF events in this sample amounts to 25%. The fractions of events for background processes LO and Compton are about 35% and 40%, respectively. The simulation was taking into account kinematic smearing in the reconstruction program used in the experiment as well as losses of tracks due to chamber inefficiencies.

The sample described above is used for several tests of selection procedures presented in this paper. The following variables were used in the classification:

- variables characterizing the event:
 - ν, y, Q^2, x_{Bj} ,
 - multiplicity of tracks,
- variables for two hadrons with highest p_T :
 - p_T, p_L — transverse and longitudinal momentum of hadron with respect to the virtual photon direction,
 - charge of the hadron,
 - $z = E_h/\nu$, where E_h is the hadron energy in the laboratory frame,
 - Φ — azimuthal angle between transverse momenta of this hadrons.

5. Results

In this section we compare results of selection of the PGF process for three classification techniques: the most commonly used selection based on cuts, the Bayes method and the neural network. From arguments discussed in the introduction one expects that the high sensitivity on the PGF process is related to $p_{T(1-2)}$ of the selected hadrons. Also the variable y is strongly correlated with the size of the expected asymmetries. These three variables were found to be most efficient for the selection based on cuts. That is why as first approach we compared different techniques in the same conditions using only tree variables mentioned above: p_{T1} , p_{T2} , y . The results on the purity and the efficiency are shown in Fig. 7. The dashed and solid lines show the Bayes method and the neural network results, respectively, and points correspond to several cuts. It was found that optimal selection for the cut method is obtained for cuts on $(p_{T1}^2 + p_{T2}^2)$ (set (i)). The values for every cut number are listed in Table I. As an example another set of cuts (ii) is also presented in Fig. 7. For the hadron with highest p_T , the cut on p_T is changing while for the other one is kept at 0.8 GeV. This selection is less efficient than the set of cuts (i). As is seen in Fig. 7 there are no significant differences when we compare the best selection based on cuts (set (i)) with the neural network and Bayes results.

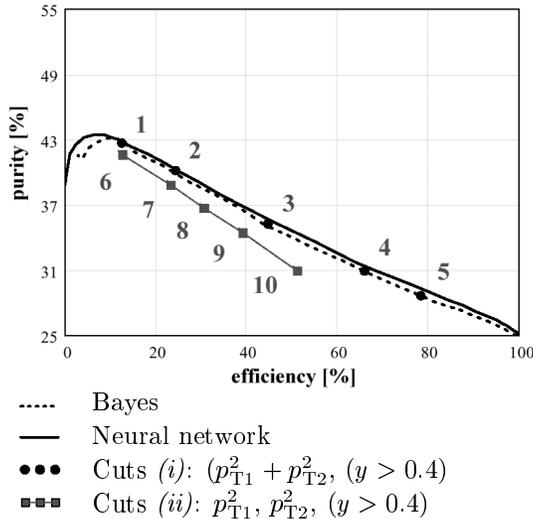


Fig. 7. Comparison of all techniques. Used variables: y , p_{T1} and p_{T2} .

Up to now we used limited set of variables. If we try to select the PGF process by handling more variables the manual searching for an optimal set of cuts costs a lot of time and does not bring improvements. In case

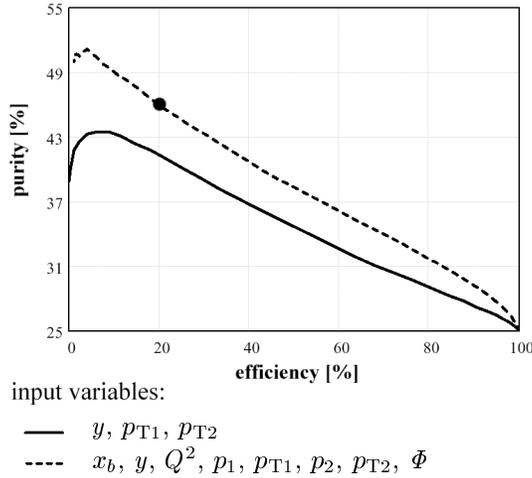


Fig. 8. Results for the neural network using limited and optimal set of variables.

TABLE I

The list of cuts

the (i) selection		the (ii) selection		
number	$p_{T1}^2 + p_{T2}^2$	number	p_{T1}	p_{T2}
1	3.2	6	1.4	0.8
2	2.5	7	1.2	0.8
3	1.9	8	1.1	0.8
4	1.5	9	1.0	0.8
5	1.3	10	0.9	0.8

of Bayes technique, when we want to use more variables, the processing of each event takes much time because of the vast training set. The situation is much better for the neural network due to simplicity of computing. For this reason we decided to use only the neural network in further tests.

It is not clear how to choose variables other than y and p_T to compose the input vector for the network because there are no obvious differences in the distributions for PGF and background processes. A lot of attempts were made to obtain the best combination.

Finally, the optimum set of variables was found as follows: x_b, y and Q^2 for variables describing the event and p_T, p_L and Φ for variables characterizing the selected hadrons. The addition of the charge of the selected hadrons and also the multiplicity and energy of all hadrons in the event do

not improve the result. The comparison of the results of the classification done by the neural network with the optimal set of variables and using only information about y and p_T is presented in Fig. 8. The purity at a given efficiency for the optimal set of variables (dashed line) is a few percent better than the results for the neural network which uses a limited set of variables.

Each point on the line which describes the result of classification done by the neural network corresponds to a certain output threshold value. Choosing a low threshold we get high efficiency but purity is low due to large contributions from background processes. The opposite situation occurs at high thresholds.

From the purity dependence on efficiency shown in Fig. 8 the choice of the optimal threshold for further physical analysis is not obvious and an additional criterion is needed. To solve this problem the error on the gluon polarization $\frac{\Delta G}{G}$ as the function of threshold is calculated using the expression given by Eq. (2.4). In this calculation the statistical error on the measured asymmetry δA and the systematic error δR on ratios R are taken into account. The error on A_1 asymmetry is neglected in the calculation as it is small compared to δA . Also no error is introduced on partonic asymmetries \hat{a}_{LL} calculated at leading order.

The results of the calculation of the error on $\frac{\Delta G}{G}$ as a function of the threshold are shown in Fig. 9 assuming different statistics of the sample and also varying the error on δR . The size of the sample which was taken into account corresponds to: (a) one year of data taking by SMC experiment and

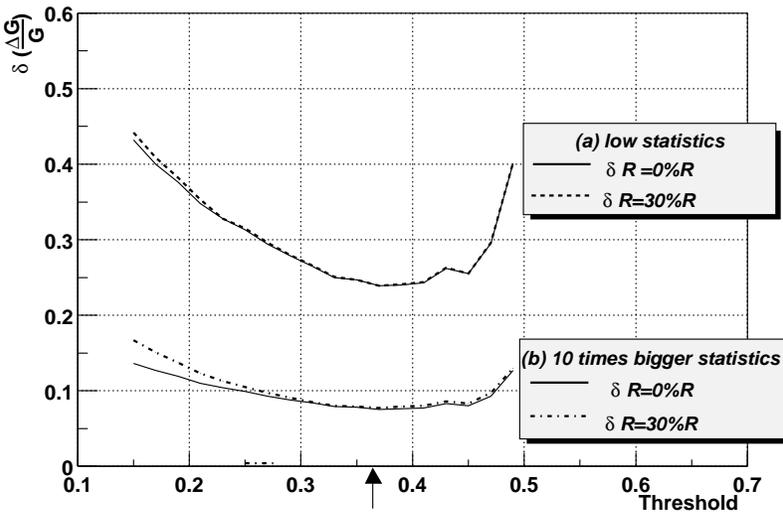


Fig. 9. The dependence $\frac{\Delta G}{G}$ on the threshold.

(b) ten times bigger statistics than (a). For both cases two possibilities of an error on δR were considered. First δR is neglected (the solid lines in Fig. 9), second the error is taken as $\delta R = 30\%R$ (the dashed and dashed-dotted lines in Fig. 9).

The observed increase of the error on $\frac{\Delta G}{G}$ at low threshold is the result of the small fraction of PGF events in the sample while at high threshold it is mainly due to loss of statistics. The optimum threshold was obtained around 0.37 as is marked by the arrow in Fig. 9.

For low statistics, the contribution from δR is negligible. For bigger statistics the systematic error δR begins to play a role at low thresholds but the optimal value of threshold does not change when we increase the systematic error δR . The minimum is getting broader but covers the same region of threshold values. This allows to select different thresholds, giving the possibility of systematics checks. The choice of the threshold at 0.37 corresponds to a purity of the sample of $\sim 45\%$ and an efficiency of $\sim 20\%$ marked as a dot in Fig. 8. For this threshold value the precision on gluon polarization determination is expected to be better than $\delta(\Delta G/G) \approx 0.3$ for SMC experiment. This can be compare to result obtain with cut selection method $\delta(\Delta G/G) \approx 0.46$. If the statistics of the sample is bigger we get $\delta(\Delta G/G) \approx 0.1$ for neural network and $\delta(\Delta G/G) \approx 0.15$ for method using cut selection.

6. Summary

We developed the neural network algorithm for the classification of PGF process. Many tests prove that this algorithm is stable and technically well performing. Also we show that neural network method is comparable to the standard analysis for three variables used by traditional cut selection and brings improvement by few percent in purity if extra variables are added. Neural network make optimal use of many corelated variables without significant penalty in computing time. The application of this method to the simulations of the SMC experiment leads to 45% purity of the selected sample which is sufficient for gluon density determination with precision better than $\delta(\Delta G/G) \approx 0.3$. It shows that the method can be successfully applied for the experimental data analysis.

REFERENCES

- [1] EMC Collaboration, J. Ashman *et al.*, *Nucl. Phys.* **B328**, 1 (1989).
- [2] E142 Collaboration, P.L. Anthony *et al.*, *Phys. Rev.* **D54**, 6620 (1996).
- [3] E143 Collaboration, K. Abe *et al.*, *Phys. Rev.* **D58**, 112003 (1998).

- [4] E154 Collaboration, K. Abe *et al.*, *Phys. Rev. Lett.* **79**, 26 (1997).
- [5] E155 Collaboration, P.L. Anthony *et al.* *Phys. Lett.* **B458**, 529 (1999).
- [6] SMC Collaboration, D. Adams *et. al.*, *Phys. Rev.* **D56**, 5330 (1997).
- [7] HERMES Collaboration, A. Airapetian *et. al.*, *Phys. Lett.* **B442**, 484 (1998).
- [8] COMPASS , G. Baum *et al.*, COMPASS: A Proposal for a Common Muon and Proton Apparatus for Structure and Spectroscopy, CERN-SPSLC-96-14.
- [9] G. Radel, Future Physics with Polarised Protons at HERA, Given at 6th International Workshop on Deep Inelastic Scattering and QCD (DIS 98), Brussels, Belgium, 4-8 Apr 1998.
- [10] PHENIX/Spin Collaboration, Proposal on Spin Physics Using the RHIC Polarized Collider, BNL-PROPOSAL-R5.
- [11] S.D. Bass, A.W. Thomas, *Prog. Part. Nucl. Phys.* **33**, 449 (1994).
- [12] SMC Collaboration, B. Adeva *et. al.*, *Phys. Lett.* **B412**, 414 (1997).
- [13] E155 Collaboration, P.L. Anthony *et al.*, *Phys. Lett.* **B493**, 19 (2000).
- [14] A. Konig, *Z. Phys.* **C18**, 63 (1983).
- [15] A. Bravar, D. von Harrach, A. Kotzinian, *Phys. Lett.* **B421**, 349 (1998).
- [16] HERMES Collaboration, A. Airapetian *et al.*, *Phys. Rev. Lett.* **84**, 2584 (2000).
- [17] A. Bravar, K. Kurek, R. Windmolders, *Comput. Phys. Commun.* **105**, 42 (1997).
- [18] T. Masters, *Practical Neural Network Recipes in C++*, Academic Press, New York 1993.
- [19] M.A. Ajzerman, E.M. Brawerman, L.I. Rozonoer, *Rozpoznawanie Obrazów. Metoda Funkcji Potencjalnych*, PWN, 1976, in Polish.
- [20] J. Hertz, A. Krogh, H.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Reading 1991.
- [21] J.H. Zurada, *Introduction to Artificial Neural Systems*, West Publishing Company, St.Paul 1992.
- [22] E. Polak, *Computational Methods in Optimization*, Academic Press, New York (1971)
- [23] M.H. Hassoun, *Fundamentals of Artificial Neural Networks*, MIT Press, Cambridge, Massachusetts 1995.
- [24] P.P. Van der Smagt, *Neural Networks* **7**, 1 (1994).
- [25] G. Ingelman, A. Edin, J. Rathsman, *Comput. Phys. Commun.* **101**, 108 (1997).
- [26] M. Gluck, E. Reya, A. Vogt, *Z. Phys.* **C53**, 127 (1992); M. Gluck, E. Reya, A. Vogt, *Z. Phys.* **C67**, 433 (1995).
- [27] T. Gehrmann, W.J. Stirling, *Phys. Rev.* **D53**, 6100 (1996).