

Heavy flavor identification using multivariate analysis at H1

Mila Pandurović, Ivanka Božović-Jelisavčić, Mihajlo Mudrinić

VINČA Institute of Nuclear Sciences, Belgrade, Serbia

Abstract. We discuss b quark identification in deep inelastic scattering of electron on proton at H1 by applying multivariate analysis method. Separation between heavy and light flavors can be further used to extract proton quark content.

Keywords: e-p DIS, b-tagging, TMVA
PACS:

INTRODUCTION

An important milestone for the study of the proton structure was the operation of HERA ep collider located at DESY, Hamburg, Germany. During its first part of operation (1992-2000) denoted as HERA I, HERA provided 820 GeV protons colliding with 27.5 GeV electrons/positrons¹ leading to the centre-of-mass energy of the collision $\sqrt{s} \sim 300$ GeV. After an upgrade HERA II (2002) proton beam energy was raised to 920 GeV ($\sqrt{s} \sim 320$ GeV). HERA ended its operation in summer 2007.

Deep Inelastic lepton-hadron scattering (DIS) at HERA is of the crucial importance for the understanding the structure of the nucleon and of dynamics of parton interaction. Measurement of proton structure functions is important test of QCD as well as an input for the proton-proton collider LHC.

Heavy quark production is an important process to study in quantum chromodynamics (QCD). It is expected that perturbative QCD (pQCD) at next-to-leading order (NLO) should give a good description of heavy flavor production in deep-inelastic scattering (DIS).

Performance of heavy quark tagging affects precision of measurement of the structure functions. Also, the physics program at LHC is deeply involved with SM searches of the Higgs boson, and in that sense proton structure functions are of the foremost importance for calculation of production rates of the related processes.

A tagging method is called inclusive when it's not relying on particular decay channel. At HERA inclusive approach has been used for determination of structure functions for heavy quarks [1,2]. For the heavy quark tagging method an inclusive approach is used, which exploits the different lifetime signature, as well as, the mass of charm and beauty flavored hadrons compared to light hadrons. The lifetime signature of charm and beauty flavored hadrons is reflected in tracks which are displaced from the ep collision point (primary vertex PV).

Events containing heavy quarks are distinguished from those containing only lights quarks using multivariate analysis techniques. The analysis is based on a sample of e⁺p Neutral Current scattering events corresponding to an integrated luminosity of 54.4 pb⁻¹ taken in the 2006 at ep center-of-mass energy of ~ 319 GeV, obtained by H1 experiment at HERA.

¹ The term "electron" will be used throughout the paper for both electrons and positrons

H1 DETECTOR

The H1 detector is one of the four experiments at HERA: two multipurpose (H1 and ZEUS) and two fixed-target experiments (HERMES and HERA-B), which were located along the storage rings. H1 is a general-purpose detector for energy and momentum measurement of charged and neutral particles. Only a short description of the H1 detector will be given here; a full description may be found in [3]. A right-handed coordinate system is employed at H1 that has its z-axis pointing in the proton beam (forward) direction and x (y) pointing in the horizontal (vertical) direction.

Charged particles are measured in the central tracking detector (CTD). The device consists of several tracking detectors. The main one is the Central Jet chamber, which consist of two cylindrical drift chambers interspersed with z-chambers to improve the z- coordinate reconstruction. CTD is situated in uniform 1.15 T magnetic field and the polar angle coverage is $20^\circ < \theta < 160^\circ$.

In order to provide better spatial resolution CTD track are linked with a hits from microvertex detector – Central silicon tracker (CST). CST consists of double-sided silicon strip detectors surrounding the beam pipe. Polar angle coverage is $30^\circ < \theta < 150^\circ$. The present analysis uses only information in $r\varphi$ plane. The resolution of transverse distance of closest approach of a track to the nominal vertex (DCA) in the $r\varphi$ plane is $33 \mu\text{m} \oplus 90 \mu\text{m} / p_t$ [GeV], where p_t represents transverse momentum of the track.

The track detectors are surrounded with the Liquid Argon calorimeter ($4^\circ < \theta < 155^\circ$) and lead-scintillating fibre calorimeter SPACAL ($153^\circ < \theta < 178^\circ$) with electromagnetic and hadronic sections. These calorimeters are used for energy and angular reconstruction of hadronic final state particles from the hadronic system and are also used to measure and identify the scattered electron.

Deep Inelastic Scattering

Quark parton model (QPM) is the theoretical model in which constituents of hadrons are considered to be non-interaction quasi-free particles (quarks). The underlying picture is the infinite momentum frame: a reference frame in which momentum of proton is large in comparison to the proton mass. A schematic Feynman graph of deep inelastic scattering process in QPM interpretation is given in Figure 1. Neutral Current scattering processes, which are considered in this paper, occur via exchange of γ or Z bosons. The incoming electron (with four-momentum k) scatters of the proton (four-momentum p) to a final state electron (four-momentum k'), via virtual photon with virtuality Q^2 , exchanged in t-channel. The kinematics of ep scattering can be described using two independent variables for fixed centre-of-mass energy. The characteristic variables used in this analysis are Q^2 and x , defined as:

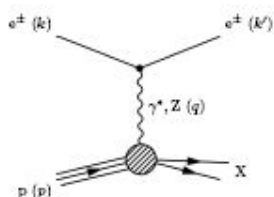


FIGURE 1. Leading order Feynman diagram for Neutral Current deep inelastic electron - proton scattering.

$$Q^2 = -q^2 = -(k - k')^2, \quad Q^2 \geq 0,$$

$$x = \frac{Q^2}{2p \cdot q}, \quad 0 \leq x \leq 1,$$

where Q^2 is virtuality of gauge boson and x is Bjorken scaling variable, which represents (in QPM) a fraction of protons momentum carried by the struck quark.

EXPERIMENTAL METHOD

Quark Flavor Separation

For the quark flavor separation we used lifetime information of heavy hadrons, as well as mass of hadrons. The lifetime of heavy hadrons will reflect in greater impact parameter compared to the light hadrons, while the mass of heavy hadrons will result in enlarged mass and multiplicity of the corresponding jet.

Impact parameter is a transverse distance of closest approach (DCA) of track to the primary vertex. If the angle between the quark axis and the line joining PV to the point of DCA is less than 90° , δ is defined as positive, otherwise it is defined as negative. Hadrons from light quarks, with a short lifetime, are viewed as coming from PV so DCA should peak at zero. For heavy hadrons long lifetimes will lead to the excess of heavy flavor events in positive values of DCA. Based on the signed impact parameter we define significance as:

$$S = \frac{\delta}{\sigma(\delta)},$$

where δ is the signed impact parameter and $\sigma(\delta)$ the absolute error of it. Significances S_1 , S_2 are defined as the significance of the tracks with the highest and second highest absolute significance. The highest and second highest transverse momentum of a tracks belonging to a jet are denoted P_{t1} and P_{t2} .

We selected events with number of tracks greater or equal 2, which have both significances of the same sign, to ensure that the tracks originate from the secondary vertex (SV). The inclusive approach using significance was used for b-tagging at HERA [1,2].

Hadrons coming from heavy quarks have large masses in comparison to hadrons coming from the light quarks. Moreover, the mass of beauty flavored hadrons are more than two times larger than charm hadrons. This characteristic will result in enlarged mass of a corresponding b-jet. Jet mass is defined as a sum of four momentum of every particle (track) that belongs to a jet. Also, transverse momentum of a track is used as an observable sensitive to lifetime information.

For separation of beauty quark events from charmed and light flavored we used a TMVA package [4], which is an ROOT [5] integrated environment for the processing and parallel evaluation of sophisticated multivariate (MV) classification methods.

All the methods are subjected to two phases: “learning” and application phase. In the learning phase a sample of variables, which are sensitive to different behavior of signal and background, serve as an input to the package. Each variable, at each step of training in the learning process, is assigned with a certain weight. The assigned weights represent the importance of variable for discrimination between signal and background i.e. weights are “knowledge” of the MV methods on how to discriminate the signal from background. Optimization of signal to background discrimination is achieved by maximizing purity-efficiency distribution. In application phase we used TMVA output variables to obtain quark fractions in proton.

RESULTS

Signal and background samples were generated with the RAPGAP program [6]: signal sample consisted of 10^4 b events; background were 10^4 c events and 10^4 light quark events (u,d,s). They are used to train TMVA package as well as to fit data and Monte Carlo samples in order to measure quark fractions.

In order to define the most optimal choice of variables sensitive to b flavor, different sets of variables are used:

Set 1: S_1

Set 2: $S_1, S_2, M_J, N_J, P_{tAv}$

Set 3: $S_1, S_2, P_{t1}, P_{t2}, P_{tAv}, M_J, N_J, P_{tlet}$

where:

- S_1 and S_2 are highest and second highest significance of tracks in jet
- M_J is mass of jet
- N_J is number of tracks in jet

- P_{tAV} is mean value of highest and second highest transverse momentum of tracks in jet
 $P_{tAV} = (P_{t1} + P_{t2})/2$
- P_{tjet} is momentum of a jet defined as $P_{tjet} = \sum_i P_{ti}$

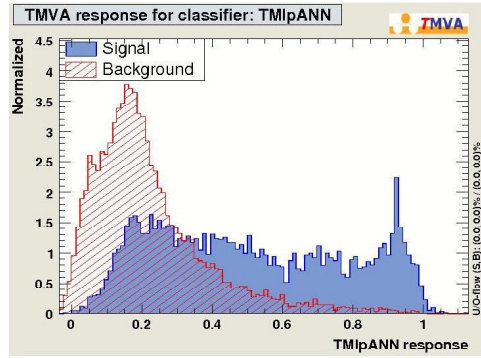


FIGURE 2. Output of TMlpANN method for signal (blue) and background (red).

Cut-off value of TMVA output variable determines level of discrimination of signal and background. The output variable for TMlpANN method is given on the Figure 2. Optimal cut-off value is obtained by maximizing the product of efficiency and purity (which corresponds to background rejection).

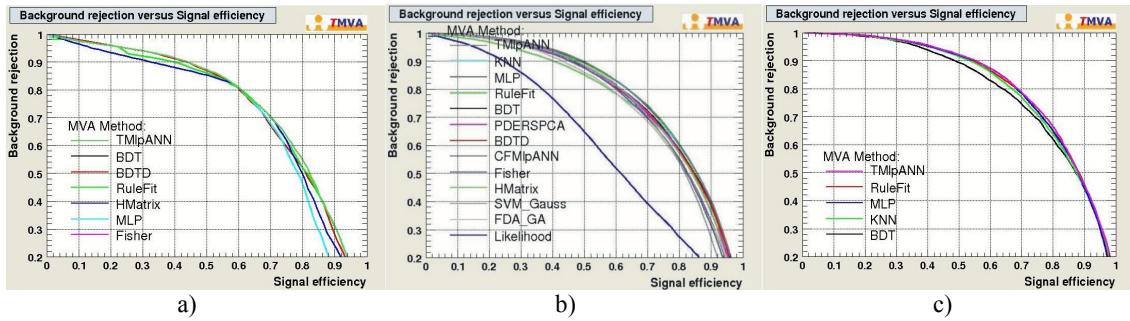


FIGURE 3. Efficiency vs. background rejection curves for different input variable sets: (a) Set1, (b) Set2, (c) Set3.

Corresponding efficiency vs. background rejection plots for the tested set of variables are given on the Figure 3. For Set1 signal efficiency of 90 % corresponds to ~ 30% of background rejection, while Set2 and Set3 exhibit improvement of approximately 10 %, 15% respectively, for the best method used. The separation of signal and background is showing overall improvement with the number of sensitive observables used. Nevertheless S1 is very powerful variable for b-tagging, even when solely used. Figure 3. is also showing the ranking of the MV methods. Also, as can be seen from Figure 3., the most of the methods are giving approximately the same response to the input variable sets.

In order to obtain the quark fractions we used output variable of the TMlpANN method obtained on data and MC samples. Data and total MC are fitted with appropriate signal (b) and background (Uds,c) samples. TMlpANN output variable for data and signal and background is shown on Figure 4.

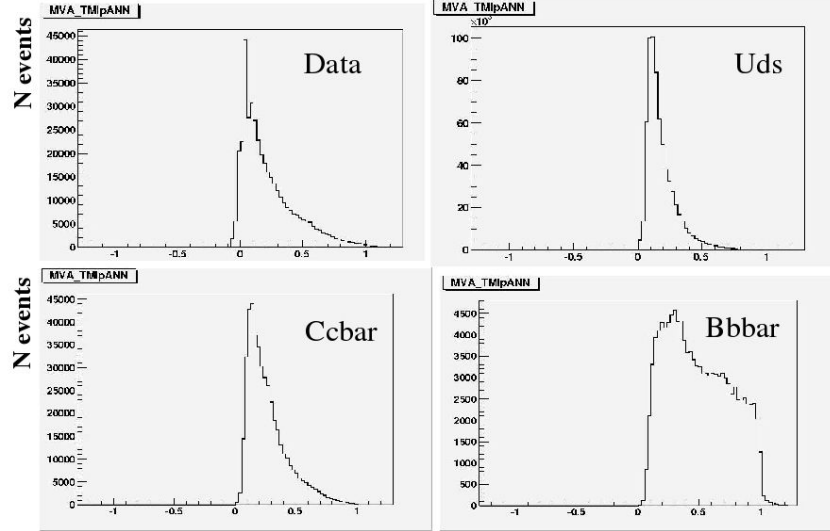


FIGURE 4. Output variable of TMlpANN method for data and uds, c, b samples.

Data and Monte Carlo samples are fitted in four Q^2 bins using Barlow fitter [7]. Obtained quark fractions are given in Table 1). Also, we tested performance of the Barlow fitter by comparing fitted MC fractions with the counted ones (red figures in Table 1).

TABLE (1). Obtained quark fractions in proton for a) Data and b) Total Monte Carlo

GeV^2	a) Data			b) Total MC		
	uds %	c%	b%	uds%	C%	b%
$6 < Q^2 < 20$	58.35 ± 0.53	40.89 ± 0.58	0.76 ± 0.13	67.94 ± 0.48	31.06 ± 0.50	1.00 ± 0.11
$20 < Q^2 < 40$	66.73 ± 0.68	32.40 ± 0.73	0.86 ± 0.17	67.16 ± 0.57	31.80 ± 0.59	1.03 ± 0.14
$40 < Q^2 < 100$	70.90 ± 0.78	28.24 ± 0.85	0.85 ± 0.23	66.38 ± 0.62	31.94 ± 0.68	1.68 ± 0.18
$Q^2 > 100$	75.97 ± 1.02	22.89 ± 1.13	1.14 ± 0.32	69.76 ± 0.73	27.83 ± 0.81	2.41 ± 0.24
				67.76	31.26	0.98
				66.95	31.95	1.10
				65.54	32.84	1.62
				68.95	28.41	2.64

As can be seen from Table 1., fractions obtained on the total Monte Carlo sample are in good agreement with the counted quark fractions, in each Q^2 bin proving the performance of the Barlow fitter. The obtained data fractions are in overall agreement with the total Monte Carlo, having in mind differences induced by the generator-level choice of parton density functions.

CONCLUSION

Multivariate classification methods can be used as a powerful tool for b tagging. Easy application of the TMVA toolkit makes it possible to find a suitable method and optimal set of parameters.

It has been shown separation of b-quark events can be improved with increasing number of sensitive observables (lifetime and mass of heavy quarks together with kinematical quantities P_{t1} , P_{t2} , P_{tAV} , P_{tjet}). Still, the highest significance is a very powerful variable for b-tagging, even when solely used.

Measured fractions of light and heavy quarks are in agreement, within less than 10 %, with the ones obtained from the Monte Carlo sample, where the differences are induced by the generator-level choice of parton density functions.

ACKNOWLEDGMENTS

We are grateful to the Ministry of Science and Technology of the Republic of Serbia for supporting this research activity within the project 'Physics and detector studies in HEP Experiments'.

REFERENCES

1. H1 Collab., A. Aktas et al., *Measurement of F_2^{cc} and F_2^{bb} at Low Q^2 and x using the H1 Vertex Detector at HERA*, Eur. Phys. J. C45 (2006) pp. 23-33.
2. H1 Collab., A. Aktas et al., *Measurement of F_2^{cc} and F_2^{bb} at High Q^2 using the H1 Vertex Detector at HERA*, Eur. Phys. J. C40 (2005) pp. 349-359.
3. H1 Collab., I Abt et al., *The H1-detector at HERA*, Nucl. Instrum. Meth. A386(1997) pp. 310 and 348.
4. TMVA homepage, (<http://tmva.sourceforge.net/>).
5. ROOT, *Data analysis framework*, (<http://root.cern.ch/drupal/>).
6. H. Jung, *"Hard diffractive scattering in high energy physics and the Monte Carlo generation RAPGAP"*, Comp. Phys. Comm. 86(1995) pp. 147.
7. R. Barlow and C. Beeston, *"Fitting using finite Monte Carlo Samples"*, Comp. Phys. Comm. 77(1993) pp.219-228.