LHCb Topological Trigger Reoptimization

Tatiana Likhomanenko, Philip Ilten, Egor Khairullin, Andrey Ustyuzhanin, Michael Williams

E-mail: `tatiana.likhomanenko@cern.ch`

Abstract. The main b-physics trigger algorithm used by the LHCb experiment is the so-called topological trigger. The topological trigger selects vertices which are a) detached from the primary proton-proton collision and b) compatible with coming from the decay of a b-hadron. In the LHC Run 1, this trigger utilized a custom boosted decision tree algorithm, selected an almost 100% pure sample of b-hadrons with a typical efficiency of 60-70%, and its output was used in about 60% of LHCb papers. This talk presents studies carried out to optimize the topological trigger for LHC Run 2. In particular, we have carried out a detailed comparison of various machine learning classifier algorithms, e.g., AdaBoost, MatrixNet and neural networks. The topological trigger algorithm is designed to select all "interesting" decays of b-hadrons, but cannot be trained on every such decay. Studies have therefore been performed to determine how to optimize the performance of the classification algorithm on decays not used in the training. These include cascading, ensembling and blending techniques. Furthermore, novel boosting techniques have been implemented that will help reduce systematic uncertainties in Run 2 measurements. We demonstrate that the reoptimized topological trigger is expected to significantly improve on the Run 1 performance for a wide range of b-hadron decays.

1. Introduction

The LHCb trigger is divided into three stages: a hardware, or level-0 (L0) stage, and two software, or high-level, stages (HLTs) [1]. The second stage of the HLT (HLT2) processes few enough events that it is possible to perform reconstruction that is very similar to what is done offline. This allows the HLT2 to use a multivariate analysis. There are many HLT2 lines dedicated to triggering on various types of events. This note provides a detailed description of the HLT2 topological lines reoptimized for LHC Run 2. Most n-body hadronic B decays ($n \geq 3$) are only triggered on efficiently in LHCb by these lines. This note also presents a new HLT scheme for LHC Run 2, which uses a multivariate analysis not only for topological lines (the previous HLT2 LHC Run 1 version is described in details in [2]).

2. HLT LHC Run 2 scheme

In the HLT1 Run 1 "1 track" line was used because we didn't have time for combinatorics and the track transverse momentum threshold was too high for an secondary vertex (SV) based approach to work. In Run 2 the HLT1 track transverse momentum threshold will be the same as the HLT2 for Run 1 so it is possible to use along with "1 track" line "2 body SV" one. Thus we consider a new HLT scheme (figure 1). The combination of displacement from primary vertex (PV) and high transverse momentum are the main characteristic of interesting physics. The HLT "1-track" line is looking for either one super high transverse momentum or high displacement track. The HLT "2 body SV" line uses a multivariate analysis to look for two tracks making a vertex. The HLT topological line is improved by more powerful multivariate analysis. It uses full reconstructed event to look for 2, 3, 4 and more tracks making a vertex.

The analysis was conducted on the following data: signal samples are simulated 13-TeV B-decays of various topologies, background sample is generic Pythia 13-TeV proton-proton collisions.

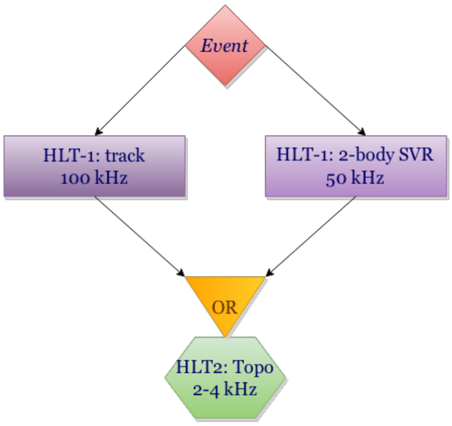

Figure 1. HLT Run 2 scheme. HLT1 consists of two lines: "1-track" and "2 body SV" with corresponding output rates 100 kHz and 50 kHz. If event passes at least one of these two lines then it is triggered by topological line with output rate 2-4 kHz.

3. Multivariate Analysis

Multivariate analysis in the HLT lines is a specific machine learning problem. Each event is represented as set of secondary vertices. These very secondary vertices are input data for classifier. Further event is triggered if at least one secondary vertex passes classifier preselection. A trigger's output rate is limited. This restriction is equivalent to the restriction on background events efficiency (FPR). For example, if an output rate is 2 kHz, then FPR is 0.2%. This can be shown using ROC curves for classifier, which is plotted for test events (figure 2).

All classifiers were trained on one half of all data and were tested on another half. For the HLT "1-track" and "2 body SV" lines all signal modes were used, while for the HLT topological line six specific modes were used in training, though all algorithms were tested on all available signal modes.

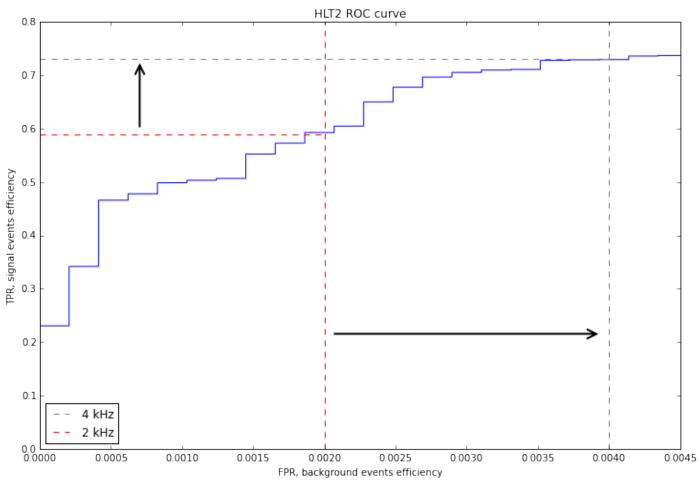

Figure 2. Trigger events ROC curve. The output rate 2 kHz corresponds to FPR 0.2%, 4 kHz — 0.4%. Thus to find signal efficiency for 2 kHz output rate we take 0.2% background efficiency and find point on the ROC curve which corresponds to this FPR.

3.1. HLT "1-track" line

Preselections and variables for the HLT "1-track" line are listed in the table 1. Only two variables are used in the line. To optimize signal efficiency and find an optimal decision boundary multivariate analysis is used. Distributions of signal and background are shown on the figure 3. MatrixNet[5] algorithm (figure 5), logistic regression (figure 4) and neural networks (figure 6) were trained to find decision boundaries. From experiments MatrixNet decision boundary is the best, but for online processing a simpler decision rule fits: some hyperbolic function can be used as the decision boundary. Efficiencies comparison of different algorithms for B-modes is shown on the figure 7.

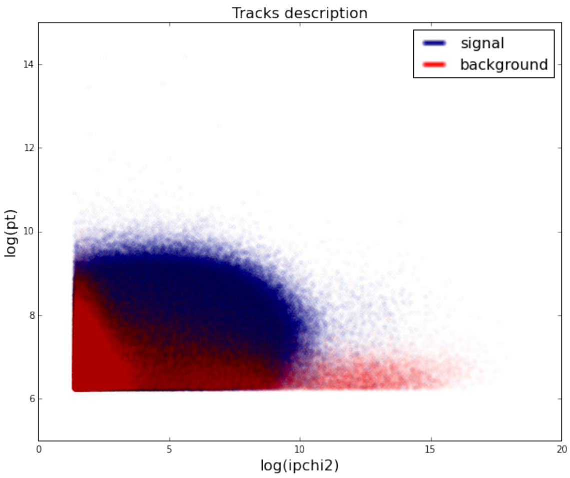

Figure 3. Track data scatters, described in two-dimensional space.

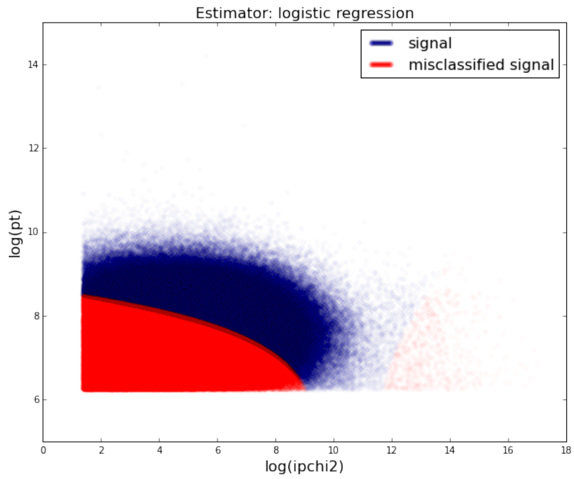

Figure 4. Decision boundary for logistic regression

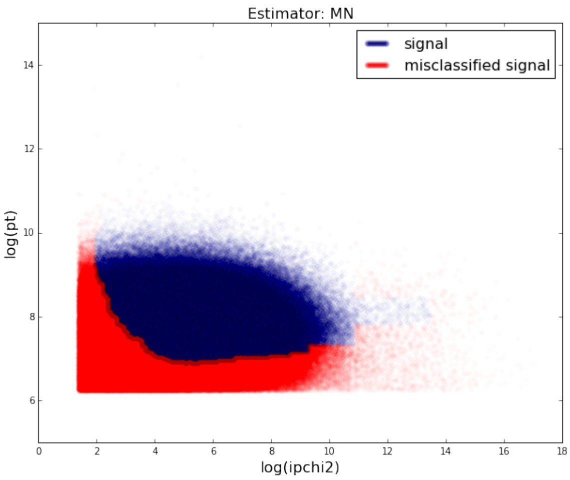

Figure 5. Decision boundary for MatrixNet algorithm.

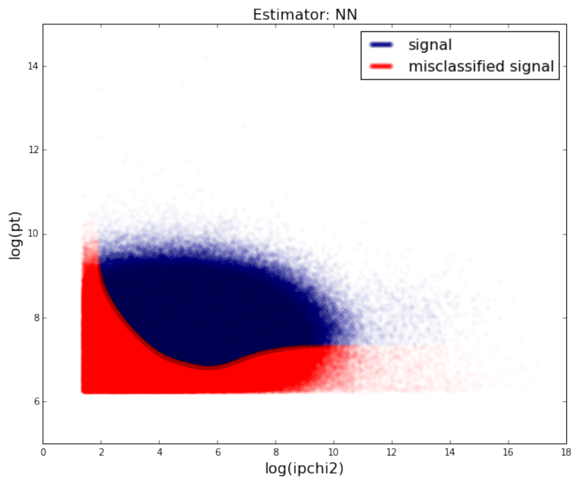

Figure 6. Decision boundary for neural networks.

Table 1. HLT "1-track" line description.

Track preselections:	
	$PT > 500$ MeV
	$IP_{\chi^2} > 4$
	$track_{\chi^2}/ndof < 3$
Analysis variables:	PT, IP_{χ^2}
Output rate:	100 kHz

3.2. HLT "2 body SV" line

Preselections and variables for the HLT "2 body SV" line are listed in the table 2. The line looks for two tracks making a vertex. In this case MatrixNet algorithm is used and several studies were conducted for it. Firstly, the possibility to remove a corrected mass cut ($mcor < 10$) and to remove the corrected mass as variable from classifier's input was investigated. These removals influence only on D-decays, but not so much. Named removals were done from systematics perspective and are expected to help in exotic searches. Secondly, to minimize the set of variables used in trigger, selection was conducted among sum, minimum of transverse momenta and just transverse momentum variables. Obtained results (efficiencies comparison) are shown on the figure 8.

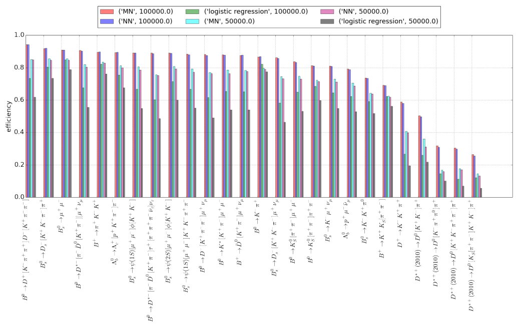

Figure 7. Efficiencies comparison for MatrixNet (MN), neural networks (NN) and logistic regression. 50 kHz and 100 kHz output rates are considered.

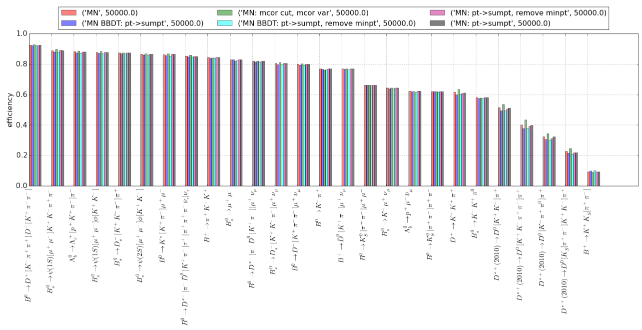

Figure 8. Efficiencies comparison for MatrixNet: with corrected mass, without it (MN), with sum of transverse momentum (sum *pt*, without min and just *pt*). These models were 'converted' to BBDT format and compared. The output rate is set to 50 kHz.

3.3. HLT topological line

The HLT2 topological lines are designed to trigger efficiently on any B decay with at least 2 charged daughters. It is designed to handle the possible omission of a daughter or daughters. To save time in the HLT2 reconstruction, only tracks with transverse momentum $PT > 200$ MeV are reconstructed. To reduce the background rate due to ghosts, all tracks are required to have a $track_{\chi^2}/ndof$ value less than 2.5. To reduce the background rate due to prompt particles, all tracks are required to have an impact parameter χ^2 value greater than 4. Due to the inclusive nature of the HLT2 topological lines, this does not mean that all of the B daughters need to

Table 2. HLT "2 body SV" line description.

Track preselections:	
	$PT > 500$ MeV
	$IP_{\chi^2} > 4$
	$track_{\chi^2}/ndof < 2.5$
SV preselections:	
	$PT > 2$ GeV
	$vertex_{\chi^2} < 10$
	$1 < mcor$ GeV
	$2 < \eta < 5$ (PV to SV)
Analysis variables:	
	sum PT, $vertex_{\chi^2}$, FD_{χ_2},
	N(tracks with $IP_{\chi^2} < 16$)
Output rate:	50 kHz

Table 3. HLT topological line description.

Track preselections:	
	$PT > 200$ MeV
	$IP_{\chi^2} > 4$
	$track_{\chi^2}/ndof < 2.5$
SV preselections:	
	$vertex_{\chi^2} < 10$
	$1 < mcor < 10$ GeV
	$2 < \eta < 5$ (PV to SV)
	N(tracks with $IP_{\chi^2} < 16) < 2$
Analysis variables:	
	n, mcor, sum PT, $vertex_{\chi^2}$,
	η, FD_{χ_2}, min PT,
	IP_{χ^2}, N(tracks),
	N(tracks with $IP_{\chi^2} < 16$)
Output rate:	2-4 kHz

satisfy these criteria. The trigger is designed to allow for the omission of one or more daughters when forming the trigger candidate. The previous topological line for Run 1 is described in [3].

In the HLT topological line Run 1 simple boosted decision tree is used [3] to define interesting secondary vertices. For Run 2 algorithm was reoptimized. Current preselections, training variables are listed in the table 3. The output rate is one of the signal efficiencies factor. The efficiency dependence on the output rate is shown on the figure 9. Thus several modes, including two training ones, significantly depend on the output rate.

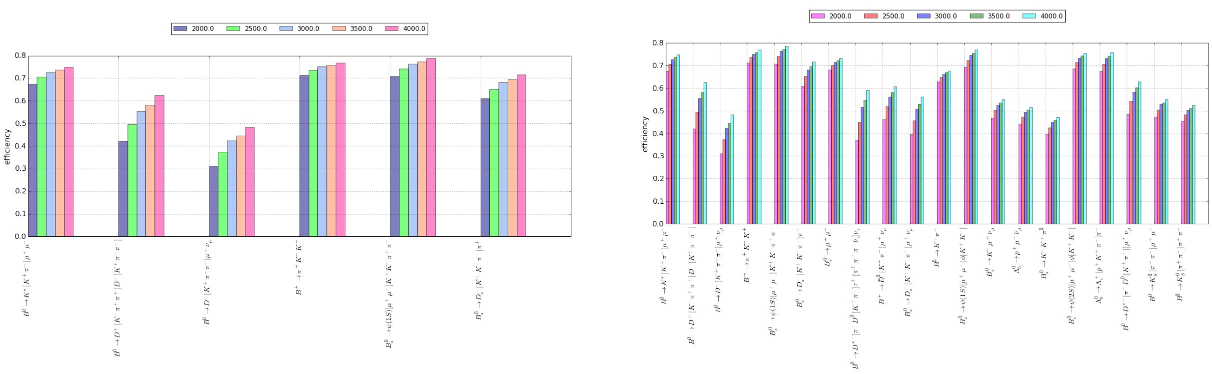

Figure 9. Signal efficiencies for training modes (left) and other available modes (right) for different output rates.

Different boosted decision trees were trained (figure 10). Also some hierarchical algorithm was conducted, so called 'blend'. Training data was divided into two parts. On the first part three MatrixNet classifiers were trained, which use only 2, 3 or 4 body decays as inputs corresponding. Then the second part was predicted by the corresponding n-bodies classifiers. These predictions are considered as new additional input variables. Resulting MatrixNet was trained on this second part using basic variables and these produced three ones. This hierarchical algorithm gives improvement for several training modes.

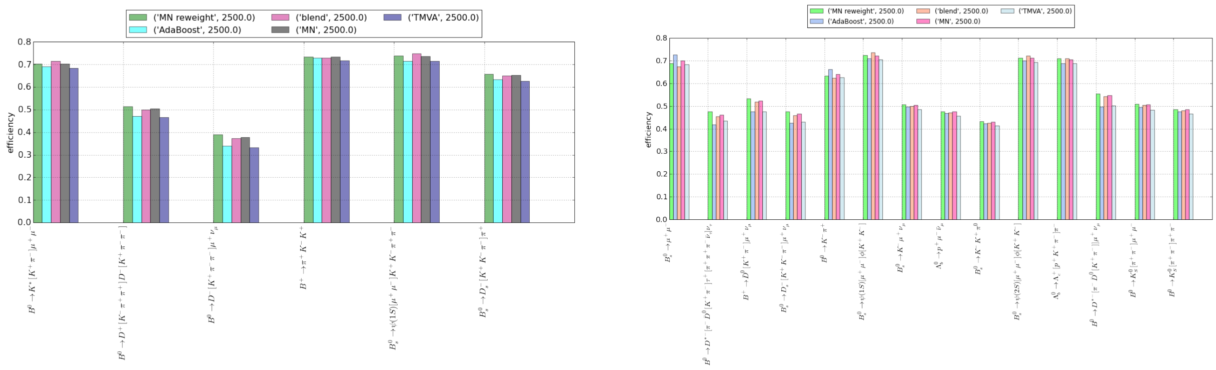

Figure 10. Comparison of different algorithms: MatrixNet (MN), MatrixNet with SV weights (MN reweight), scikit-learn AdaBoost implementation (AdaBoost), TMVA AdaBoost implementation (TMVA) and some hierarchical algorithm (blend).

From experiments MatrixNet (MN) model was chosen. Its efficiency was compared with the previous Run 1 algorithm. For these six training modes we get significantly efficiency improvement: 15-60% for 2.5 kHz output rate and 50-80% for 4 kHz (table 4).

Table 4. Ratio of Run-2 over Run-1 for HLT2/HLT1 efficiencies. Note that the denominator is reconstructible with $PT(B) > 2$ GeV, $\tau(B) > 0.2$ ps.

mode	2.5 kHz	4. kHz
$B^0 \to K^*[K^+\pi^-]\mu^+\mu^-$	1.64	1.72
$B^+ \to \pi^+K^-K^+$	1.59	1.65
$B_s^0 \to D_s^-[K^+K^-\pi^-]\mu^+\nu_\mu$	1.14	1.47
$B_s^0 \to \psi(1S)[\mu^+\mu^-]K^+K^-\pi^+\pi^-$	1.62	1.71
$B_s^0 \to D_s^-[K^+K^-\pi^-]\pi^+$	1.46	1.52
$B^0 \to D^+[K^-\pi^+\pi^+]D^-[K^+\pi^-\pi^-]$	1.40	1.86

4. Online processing

Boosted decision trees are not appropriate for online processing events in triggers due to their low speed. Two approaches exist to overcome this restrictions. First one is so-called bonsai boosted decision format[4] (shortly BBDT). Being specific BDT, MatrixNet can be converted to this format. The second approach is post-prunning: basic MatrixNet classifier includes several thousands of trees, and post-prunning procedure reduces this amount to couple of hundreds. This results in significant speedup of prediction rate. Both ways reduce MatrixNet signal efficiencies compared to basic one. In case of BBDT we are also limited in the size of BBDT lookup table in the RAM, this also reduces efficiencies. Different BBDT versions for MatrixNet were tried and compared to post-prunning to find the optimal solution for online processing (see figure 11).

Interesting to notice that the preference between BBDT and post-prunning depends on the chosen output rate. It is clear on the figure 12, that ROCs order depends on the background efficiency (or the output rate). Another research for Run 2 triggers is connected to timing comparison of BBDT and post-prunning, which is in progress at this moment.

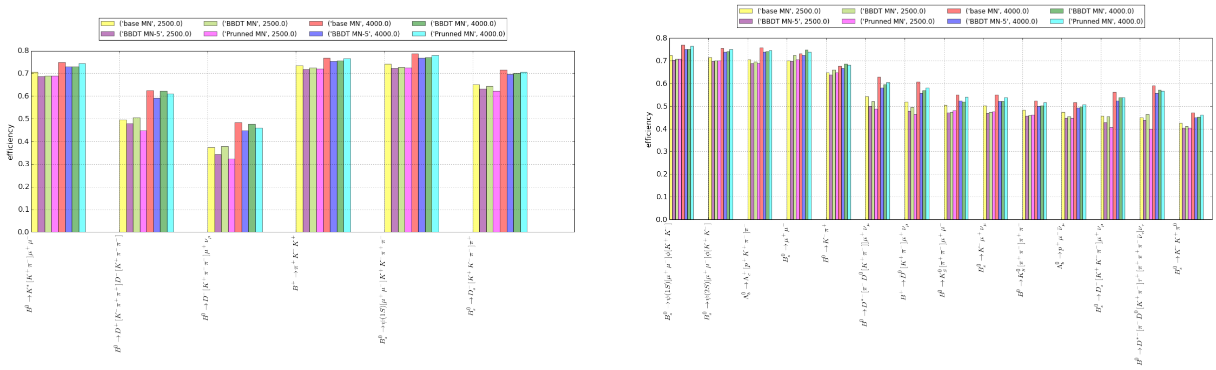

Figure 11. Comparison of basic MatrixNet (base MN), BBDT format (BBDT MN) and post-prunned MatrixNet (Prunned MN)

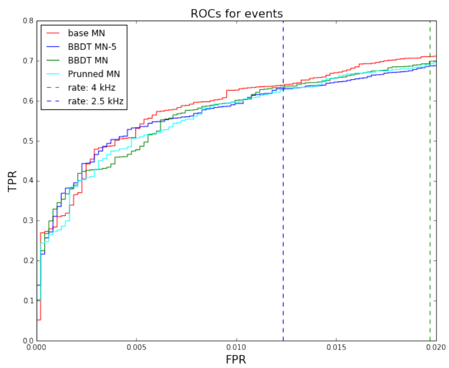

Figure 12. The best model among BBDT and post-prunning depends on background efficiency.

5. Conclusion

LHCb topological trigger was successfully reoptimized for Run 2: 15-60% efficiency improvement was obtained for 2.5 kHz output rate and 50-80% for 4 kHz. Presented HLT scheme will be applied in Run 2.

References

[1] The LHCb Collaboration. LHCb Trigger System Technical Design Report. Technical report, 2003. CERN-LHCC-2003-031.
[2] The HLT2 Topological Lines, 2011. LHCb-PUB-2011-002.
[3] The HLT inclusive B triggers, 2011. LHCb-PUB-2011-016.
[4] Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree., 2012. arXiv:1210.6861v1
[5] A. Gulin, I. Kuralenok, and D. Pavlov, Winning the transfer learning track of Yahoo's Learning to Rank Challenge with YetiRank, JMLR: Workshop and Conference Proceedings 14 (2011) 63 (Yandex MatrixNet).