Machine-Learning-based global particle-identification algorithms at the LHCb experiment

Denis Derkach[1,2], Mikhail Hushchyn[1,2,3], Tatiana Likhomanenko[2,4], Alex Rogozhnikov[2], Nikita Kazeev[1,2], Victoria Chekalina[1,2], Radoslav Neychev[1,2], Stanislav Kirillov[2], Fedor Ratnikov[1,2]
on behalf of the LHCb collaboration

[1] National Research University — Higher School of Economics, Moscow, Russia
[2] Yandex School of Data Analysis, Moscow, Russia
[3] Moscow Institute of Physics and Technology, Moscow, Russia
[4] National Research Center Kurchatov Institute, Moscow, Russia

E-mail: `Denis.Derkach@cern.ch`

Abstract. One of the most important aspects of data analysis at the LHC experiments is the particle identification (PID). In LHCb, several different sub-detectors provide PID information: two Ring Imaging Cherenkov (RICH) detectors, the hadronic and electromagnetic calorimeters, and the muon chambers. To improve charged particle identification, we have developed models based on deep learning and gradient boosting. The new approaches, tested on simulated samples, provide higher identification performances than the current solution for all charged particle types. It is also desirable to achieve a flat dependency of efficiencies from spectator variables such as particle momentum, in order to reduce systematic uncertainties in the physics results. For this purpose, models that improve the flatness property for efficiencies have also been developed. This paper presents this new approach and its performance.

1. Introduction

Particle identification (PID) algorithms play a crucial part in any high-energy physics analysis. A higher performance algorithm leads to a better background rejection and thus more precise results. In addition, an algorithm is required to work with approximately the same efficiency in the full available phase space to provide good discrimination for various analyses.

PID at the LHCb experiment relies on several subsystems [1]. Two Ring Imaging Cherenkov (RICHs) sub-detectors provide charged hadron identification over a wide momentum range, from 2 to 100 GeV/c. Muons are identified mainly thanks to the muon chambers, while electron and photon identification is assured by the calorimeters. The information can be combined by simply computing Log Likelihoods (LL) separately for each sub-detector and combining them according to the mathematical definition [2], or by using a more elaborated method like an artificial neural network (see details in Section 3). The latter outperforms the LL approach and is used as a baseline in our studies. The above particle identification approaches are used by most of the analyses completed by LHCb collaboration.

In this paper, new models for particle identification that efficiently combine the diverse information available from all sub-detectors using advanced machine learning approaches are presented. The proposed models are based on Deep artificial Neural Network (DNN)

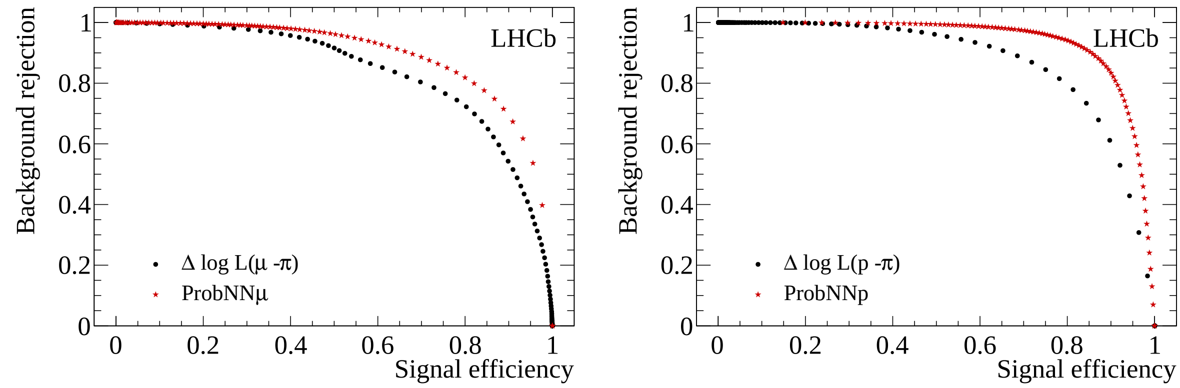

Figure 1. Background misidentification rates versus muon (left) and proton (right) identification efficiency, as measured in the $\Sigma^+ \to p\mu^+\mu^-$ decay study. The variables $\Delta\mathcal{L}(X-\pi)$ (black) and ProbNN (red), are compared for $5-10$ GeV/c muons and $5-50$ GeV/c protons, using data sidebands for backgrounds and simulated samples for the signal. The data sample used corresponds to 2012 sample collected at center-of-mass energy 8 GeV.

and gradient boosting over (ordinary and oblivious) decision trees. An additional class of algorithms providing flat efficiencies along several spectator observables (momentum, transverse momentum, pseudorapidity, number of particles in the event) is explored.

2. Problem Statement

The problem consists in identifying the charged particle type associated with a given track. There are five relevant particle species, namely, electron, muon, pion, kaon, proton, and ghost track (charged tracks that do not correspond to a real particle which passed through the detector) making a total of six hypotheses. Therefore, this is a multiclass classification problem. The information from RICHs, the electromagnetic and hadronic calorimeters and muon chambers are combined together with information provided by the tracking system. Apart from pre-aggregated into likelihood like observables subdetector responses [3], we also use track geometry variables and different detector flags. In addition to this, we used the muon identification [4] and calorimeter information about neutral clusters [5], which proved to be very useful to suppress fake tracks. In this paper, for the new methods we present only the result obtained with fully simulated events [6] used both for training and testing.

3. The current solution

The first machine learning algorithm used for the PID in LHCb is a fully-connected neural network (multilayer perceptron) with one hidden-layer implemented using the TMVA package [7]. The model, called ProbNN, was trained separately for each particle in the binary one-vs-rest classification mode, thus creating several separate models. The misidentification rates versus efficiency curves for the Log Likelihood, $\Delta\mathcal{L}(X-\pi)$, and ProbNN are shown in Figure 1. The improvement due to machine learning is clearly visible for both muons and protons.

4. New models

Two classes of algorithms are considered: first, 'non-uniform' algorithms, which have similar training target compared to the existing ProbNN; second, 'uniform' algorithms, which in addition is trained to have flat efficiencies along chosen kinematic variables.

Table 1. Relative increase of the 1-AUC scores for different particles species and ghosts (lower is better). The statistical uncertainty is lower than 1%.

Model	ghost	electron	muon	pion	kaon	proton
DNN	−29%	−41%	−52%	−37%	−20%	−17%
XGboost	−24%	−37%	−50%	−34%	−18%	−15%
CatBoost	−30%	−43%	−54%	−37%	−20%	−18%
flat 4d	−21%	−4%	−13%	−20%	+10%	+25%

We developed new PID models based on deep neural networks, from the `keras` library [8], and boosted decision trees, from the `CatBoost` [9] and `XGBoost` [10] libraries. Along with a general increase of information provided to the classifier, we used `XGBoost` library in the multiclass mode, which brought some additional improvement. To compare the algorithms' performances, we used the Area Under Receiver Operating Characteristic Curve (ROC AUC) for one-vs-rest classification (6 numbers in total). As can be seen from Table 1, the new models give encouraging results in comparison to the baseline. All numbers in Table 1 were found statistically siginficant using several methods proposed in [11, 12], paired t-test on cross-validation sample (see, for example [13]).

An important improvement that may lead to reduction of systematic uncertainties of physics annalyses is the construction of the PID model that does not depend on kinematic observables of the track. In order to flatten the efficiency dependence, we used special training procedure. This procedure effectively takes into account non-flatness of the output distribution by using the modified loss function as described in [14]. The corresponding loss function in this case looks like:

$$\mathcal{L} = \mathcal{L}_{ExpLoss} + \alpha \mathcal{L}_{FL}, \tag{1}$$

where $\mathcal{L}_{ExpLoss}$ corresponds to the classification loss function, the $\mathcal{L}_{FL}$ corresponds to the uniformity loss, and α is a parameter to control the trade-off between classification quality and uniformity. The $\mathcal{L}_{FL}$ is taken to be similar to the Cramer-von Mises measure:

$$\mathcal{L}_{FL} =< \int (F_{global}(s) - F_{local}(s))^2 \, ds >, \tag{2}$$

where $F_{global}(s)$ and $F_{local}(s)$ is a classifier predictions cumulative distribution function for the full sample or in a given interval, respectively. In order to reduce the efficiency dependency on the kinematic properties of the track we use the linear combination of flatness losses for momentum, transverse momentum, number of tracks in event and pseudorapidity:

$$\mathcal{L}_{FL_{4d}} = \mathcal{L}_{FL_p} + \mathcal{L}_{FL_{p_T}} + \mathcal{L}_{FL_{nTracks}} + \mathcal{L}_{\eta}. \tag{3}$$

This loss function is implemented using the `Decision Train` package [15], which uses fast oblivious decision trees for model construction. The results of flatness boosting approach are shown in Table 1 and in Fig. 2 with the label "flat 4d". The comparison of efficiency dependence on $1/p_T$ shows a significant improvement in flatness for each particle type. The lower performance of this model is the result of the trade-off between quality and flatness controlled by parameter α in Eq. 2. In these proceedings, we choose values of α to be 10 for track momentum, track transverse momentum components of the loss function and 6 for pseudorapidity and number of tracks components. The choice of α is motivated by the optimal trade-off between flatness and performance and is optimised using cross-validation sample. The final choice of the model depends on the particular physics analysis and should be made after

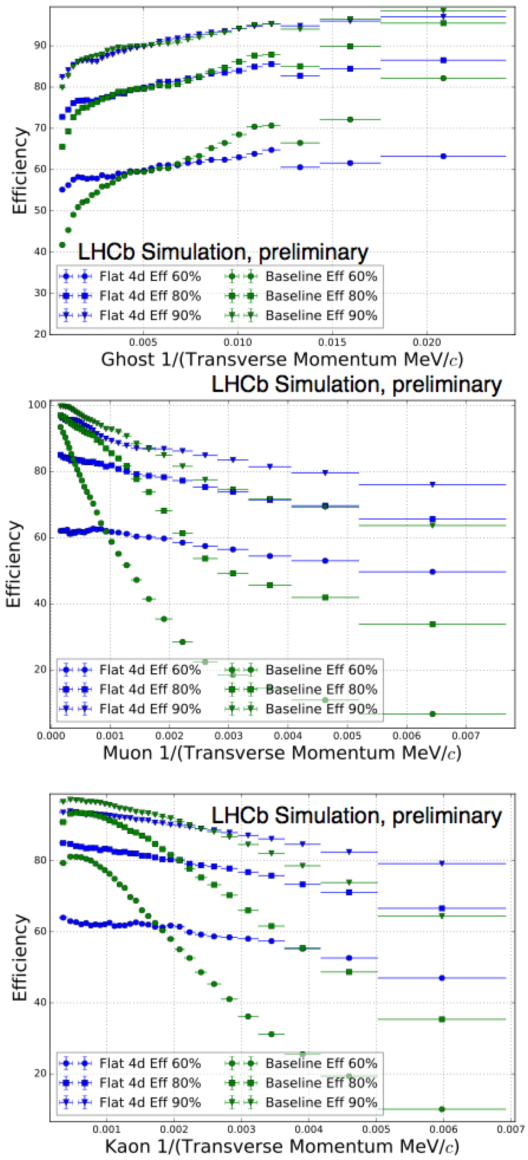

Figure 2. Efficiency of the "Flat 4d" model as a function of the inverse of the transverse momentum for each particle type, for different global efficiency cuts.

evaluation of the main sources of systematic uncertainties. Other approaches to flatten the efficiencies are also available for neural networks [16, 17].

5. Conclusions

Several new models based on state-of-the-art machine learning methods are developed and shown encouraging results in comparison to the baseline. Models drastically improving the flatness of the output with the respect to the spectator variables are built. The corresponding loss of overall efficiency is show to be small.

Acknowledgments

The research leading to these results has received funding from Russian Science Foundation under grant agreement n° 17-72-20127.

References

[1] Ignacio B *et al.* (LHCb collaboration) 2013 LHCb-TDR-014
[2] Patrignani C *et al.* (Particle Data Group) 2016 *Chin. Phys.* **C40** 100001
[3] Powell A 2011 5 p URL `https://cds.cern.ch/record/1322666`
[4] Archilli F *et al.* 2013 *JINST* **8** P10020 (*Preprint* `1306.0249`)
[5] Deschamps O, Machefert F P, Schune M H, Pakhlova G and Belyaev I 2003 Photon and neutral pion reconstruction Tech. Rep. LHCb-2003-091 CERN Geneva URL `https://cds.cern.ch/record/691634`
[6] Clemencic M *et al.* 2011 *J. Phys. Conf. Ser.* **331** 032023
[7] Hocker A *et al.* 2007 *PoS* **ACAT** 040 (*Preprint* `physics/0703039`)
[8] Chollet F *et al.* 2015 Keras `https://github.com/fchollet/keras`
[9] Dorogush A V, Gulin A, Gusev G, Kazeev N, Prokhorenkova L O and Vorobev A 2017 **abs/1706.09516** URL `http://arxiv.org/abs/1706.09516`
[10] Chen T and Guestrin C 2016 **abs/1603.02754** URL `http://arxiv.org/abs/1603.02754`
[11] DeLong E R, DeLong D M and Clarke-Pearson D L 1988 *Biometrics* **44** 837–845 ISSN 0006341X, 15410420 URL `http://www.jstor.org/stable/2531595`
[12] Hanley J A and McNeil B J 1983 *Radiology* **148** 839–843 pMID: 6878708 (*Preprint* `https://doi.org/10.1148/radiology.148.3.6878708`) URL `https://doi.org/10.1148/radiology.148.3.6878708`
[13] Demsar J 2006 *Journal of Machine Learning Research* **7** 1–30
[14] Rogozhnikov A, Bukva A, Gligorov V V, Ustyuzhanin A and Williams M 2015 *JINST* **10** T03002 (*Preprint* `1410.4140`)
[15] Rogozhnikov A 2015 Decision train classifier `http://arogozhnikov.github.io/` [Online; accessed 03-November-2017]
[16] Shimmin C, Sadowski P, Baldi P, Weik E, Whiteson D, Goul E and Sgaard A 2017 *Phys. Rev.* **D96** 074034 (*Preprint* `1703.03507`)
[17] Louppe G, Kagan M and Cranmer K 2016 (*Preprint* `1611.01046`)