Robust Neural Particle Identification Models

Aziz Temirkhanov[1]**, Artem Ryzhikov**[1]**, Denis Derkach**[1]**,
Mikhail Hushchyn**[1]**, Nikita Kazeev**[1] **and Sergei Mokhnenko**[1]
on behalf of LHCb collaboration

[1] HSE University, 20 Myasnitskaya st., Moscow 101000, Russia

E-mail: `atemirkhanov@hse.ru, aryzhikov@hse.ru`

Abstract. The volume of data processed by the Large Hadron Collider experiments demands sophisticated selection rules typically based on machine learning algorithms. One of the shortcomings of these approaches is their profound sensitivity to the biases in training samples. In the case of particle identification (PID), this might lead to degradation of the efficiency for some decays not present in the training dataset due to differences in input kinematic distributions. In this talk, we propose a method based on the Common Specific Decomposition that takes into account individual decays and possible misshapes in the training data by disentangling common and decay specific components of the input feature set. We show that the proposed approach reduces the rate of efficiency degradation for the PID algorithms for the decays reconstructed in the LHCb detector.

1. Introduction

Particle identification (PID) plays a crucial part in many high-energy physics analysis. A higher performance PID algorithm leads to a better background rejection and thus more precise results. In addition, the algorithm is required to work with high efficiency in the entire available wide range of signal topologies and kinematics and provide good discrimination for various analyses. Machine learning (ML) algorithms have gradually become the baseline approach for this task [1]. One large family of such algorithms is neural networks.

The PID algorithms of the LHCb experiment relies on several sub-detector systems [2]. Compared to other LHC experiments, LHCb has more information for hadron identification, particularly the separation between charged hadrons, provided by the Ring Imaging Cherenkov (RICH) subdetector. The muon identification is provided by the muon chambers, while the responsibility of calorimeters is mainly leptons and photons. The LHCb experiment employs several machine learning solutions [1] to aggregate the information for the PID. These solutions proved to have overall high efficiency over a large volume of the phase space. In this paper, we address one of the outstanding issues that can occur during the application of machine learning algorithms in the real-life scenario: algorithm's efficiency degradation in case the testing scenario is significantly different from training.

The paper is organized as follows. The problem formulation and description of current LHCb PID algorithms are given in Section 2. Section 3 provides a description of the Common Specific Decomposition algorithm used for the robust PID models. Finally, the list of training and testing samples, and the results for the different PID models are presented in Section 4.

2. Problem statement

The LHCb detector is a single-arm forward spectrometer covering the pseudorapidity range $2 < \eta < 5$, described in detail in Ref. [2]. Identification of various final state particles is performed by combining together the information from the LHCb subdetectors, namely from ring-imaging Cherenkov detectors (RICH), the electromagnetic and hadronic calorimeters, muon chambers and tracking system, as demonstrated in Fig. 1. Apart from the physics motivated likelihood observables based on observable subdetector responses [3], track geometry variables and different detector flags are also used. In addition, the muon identification [4] and calorimeter information about neutral clusters [5] are also used.
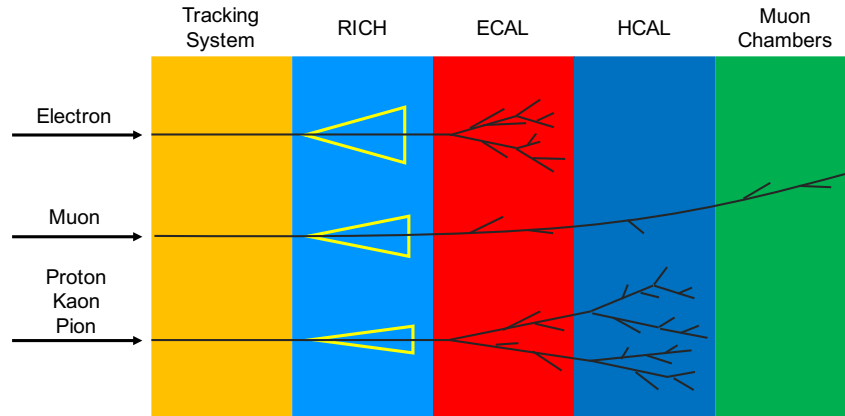

Figure 1: Illustration of different particle type responses in the LHCb systems. Figure from [6]

The objective of the PID algorithm is to identify the charged particle type associated with a given track. In the LHCb experiment, there are five relevant particle species: electron, muon, pion, kaon, and proton. A sixth hypothesis corresponds to ghost tracks, which do not correspond to a real particle that passed through the detector.
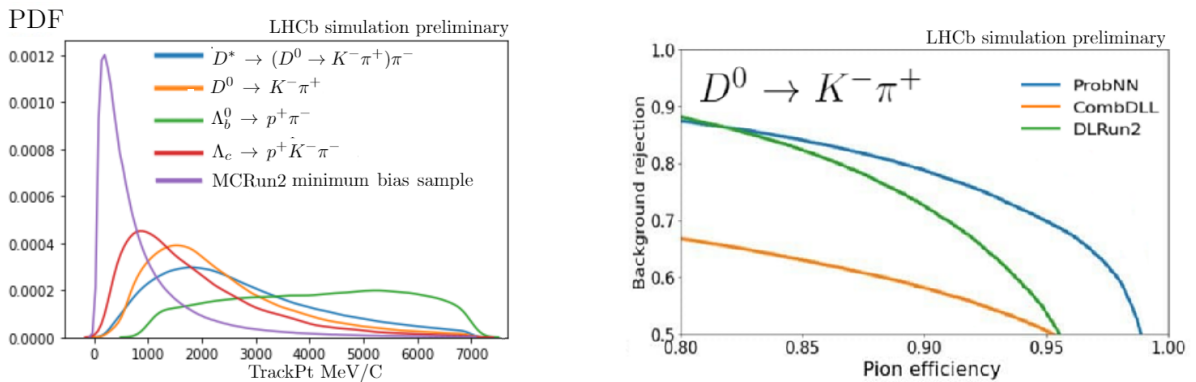

Figure 2: Distributions of transverse momentum for different decays in the simulated samples (left) and performance of the PID models on $D^0 \rightarrow K^- \pi^+$ decay for the pion-vs-all task (right).

The classification models for this task are trained on simulated events generated for the LHCb data taking condition for 2015-2018 data taking period with minimum bias trigger selection. In this study, we also use dedicated simulated samples of correctly reconstructed decays as a source of correctly reconstructed particles of a certain type. Particles in these decays might have distribution different from the training sample as demonstrated in Fig. 2. Such differences may

cause a quality reduction of the classification models in this regions. The aim of this research is to make PID algorithms [1] robust to differences in the kinematic distributions between the events used for training and events the algorithm faces in real use. The research is focused on neural networks only.

In this work, we compare our PID algorithm with the following three models: **ProbNN**, **CombDLL**, and **DLRun2**. ProbNN is a neural network with a single hidden layer. CombDLL is a combination of differential log-likelihoods from the detector's subsytems. DLRun2 is a deep neural network with 3 wide hidden layers. Neural-based models are trained on minimum bias sample. The performance of the neural-based solutions is approximately the same when tested on the minimum bias sample.

We notice that quality of the algorithms might deteriorate in some regions. The performance of the classifiers is tested with a sample of $D^0 \to K^- \pi^+$ signal candidates (Fig 2). As can be seen from the Figure, DLRun2 algorithm's quality is lower than its counterparts.

3. CSD algorithm

To address the quality degradation, we use domain adaptation techniques [7]. The idea laying behind the method is as follows: assume that there exist common features, whose correlation with the target is preserved in all decays, and decay-specific features, whose correlation differs from decays to decay (Fig 3). A classifier that relies solely on common features is robust to domain shifts.

To achieve that, CSD algorithm use a weighted combination of these three terms in training objective: orthonormality regularizes and two cross-entropy losses between labels and distributions computed from the common and domain-specific features.

This approach allows to compute low-rank decomposition along with a training networks parameters in a single layer. Hence, we replace final classification layer of DLRun2 model by CSD layer [7] as shown in Fig. 3.
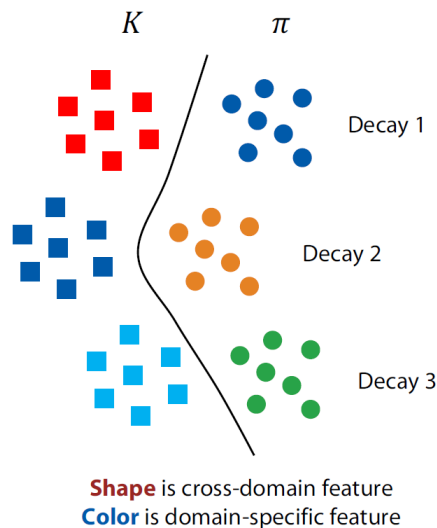

Figure 3: The CSD assumption.

4. Data and results

For the experiment we use simulated minimum bias sample and pions from several decay channels: $D^0 \to K^- \pi^+$, $D^* \to (D^0 \to K^- \pi^+)\pi^-$, $\Lambda_c \to p^+ K^- \pi^-$, and $\Lambda_b^0 \to p^+ \pi^-$. We train

CSD model similarly to DLRun2 model on minimum bias sample. Then, we test every model on simulated run2 sample and each decay channel. For the CSD model an additional round of training with mixture of minimum bias sample and subset of listed above decays has been done in order to teach model to generalize in pion domain.

In addition, we design a nearest neighbors test in the following way. For a correctly reconstructed particle in the decay chain we choose the kinematically closest particle of minimum bias sample for the different species. We thus mimic the misreconstruction real-life scenario.

Table 1 contains ROC AOC (i.e. Area over ROC curve) increment with respect to DLRun2 (uncertainty does not exceed 0.0008 as obtained by a bootstrap procedure). The result for nearest neighbors test is shown in Fig. 4.

Model	Ghost	Electron	Muon	Pion	Kaon	Proton
DLRun2	100%	100%	100%	100%	100%	100%
ProbNN	+41%	+14%	+418%	+51%	20%	+21%
CombDLL	–	+71%	+389%	+979%	+89%	+98%
CSD	**-11%**	**-42%**	**-47%**	**-0.21%**	**-10%**	**-11%**

Table 1: ROC AOCs change for different PID models for paricle-vs-all task. DLRun2 AOC is taken as 100%.
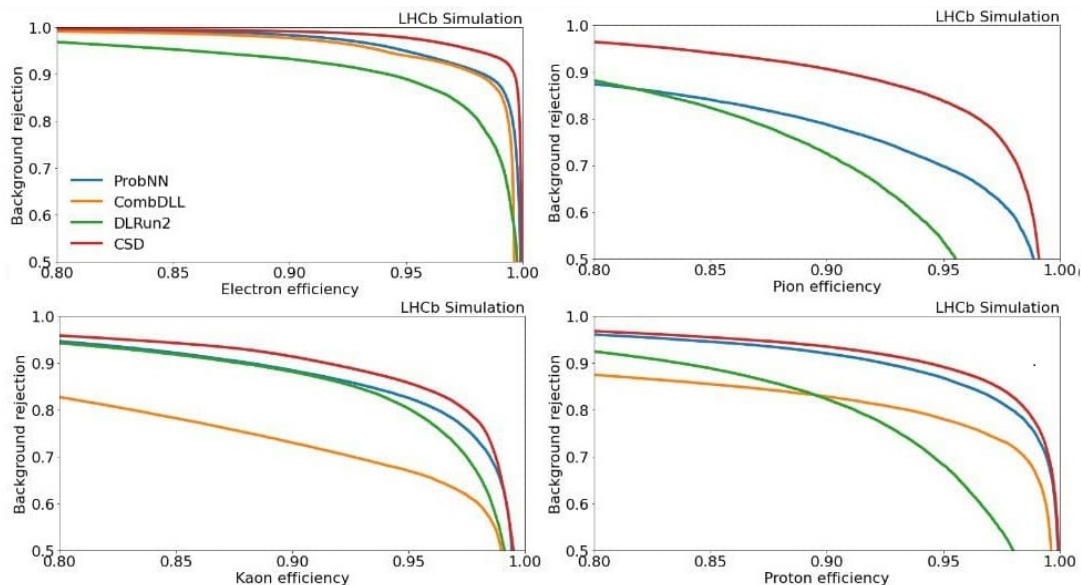

Figure 4: ROC AUCs for nearest neighbours test, the background sample is constructed of the best matching candidates as described in the text.

The quality on those 4 decays itself rises significantly, as it demonstrated in Table 2.

In addition to previous tests, we cross-check that the obtained solution has some generalization power. In order to prove this, we take two decays containing muons: $B^+ \to K^+(J/\psi \to \mu^+\mu^-)$ and $B^+ \to \mu^+\mu^-(K^* \to K^+\pi_0)$ and compare CSD and DLRun2 quality. ROC AUC error rate for proposed method almost halved. Note, that models has never seen muon decays outside a minimum bias sample before. Table 3 illustrates that model can generalize well over different domains.

4

Model	$D^* \to (D^0 \to K^- \pi^+)\pi^-$	$\Lambda_c \to p^+ K^- \pi^-$	$\Lambda_b^0 \to p^+ \pi^-$	$D^0 \to K^- \pi^+$
DLRun2	100%	100%	100%	100%
CSD	**-38%**	**-71%**	**-50%**	**-15%**

Table 2: ROC AOC for the pionic decays unseen by the CSD algorithm during training. DLRun2 AOC is taken as 100%.

Model	$B^+ \to K^+(J/\psi \to \mu^+\mu^-)$	$B^+ \to \mu^+\mu^-(K^* \to K^+\pi_0)$
DLRun2	100%	100%
CSD	**-97%**	**-98%**

Table 3: ROC AOC on muonic decays. DLRun2 AOC is taken as 100%.

5. Conclusion

The degradation of machine learning solutions quality for the PID problem in the specific parts of the phase space is addressed. The CSD algorithm based on the separation of common and domain-specific features has shown promising results. The algorithm is able to select common features even for the decays that are not present in the original domains. The obtained algorithm shows higher stability with respect to previously presented thus giving substantial increase of solution's quality for particular case. The method thus can be used in many high-energy physics applications.

References

[1] Derkach D *et al.*, *Machine-Learning-based global particle-identification algorithms at the LHCb experiment*, J. Phys.: Conf. Ser. **1085** 042038, 2018.

[2] The LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** S08005, 2008.

[3] The LHCb RICH group, *Performance of the LHCb RICH detector at the LHC*, Eur. Phys. J. C **73** 2431, 2013.

[4] Archilli F *et al.*, *JINST* **8** P10020 (*Preprint* 1306.0249), 2013.

[5] Deschamps O, Machefert F P, Schune M H, Pakhlova G and Belyaev I, *Photon and neutral pion reconstruction*, Tech. Rep. LHCb-2003-091 CERN Geneva `https://cds.cern.ch/record/691634`, 2003.

[6] D. Derkach, M. Hushchyn, and N. Kazeev, "Machine Learning based Global Particle Identification Algorithms at the LHCb Experiment," in *European Physical Journal Web of Conferences*, ser. European Physical Journal Web of Conferences, vol. 214, Jul. 2019, 06011, p. 06011. DOI: `10.1051/epjconf/201921406011`.

[7] Vihari Piratla, Praneeth Netrapalli, Sunita Sarawag, *Efficient domain generalization via common-specific low-rank decomposition*, `https://arxiv.org/abs/2003.12815`, 2020.

[8] P. Kostenetskiy, R. Chulkevich, and V. Kozyrev, "HPC Resources of the Higher School of Economics," *Journal of Physics: Conference Series.*, vol. 1740, no. 1, p. 012050, 2021. DOI: `10.1088/1742-6596/1740/1/012050`.