Machine Learning based Global Particle Identification Algorithms at the LHCb Experiment

23rd International Conference on Computing in High Energy and Nuclear Physics

Denis Derkach\textsuperscript{1}, Mikhail Hushchyn\textsuperscript{1,2}, Nikita Kazeev\textsuperscript{1,2,3}
on behalf of the LHCb collaboration

\textsuperscript{1}National Research University Higher School of Economics
\textsuperscript{2}Yandex School of Data Analysis
\textsuperscript{3}Università degli Studi Roma “La Sapienza”

Introduction

Particle identification (PID) plays a crucial role in LHCb analyses. The LHCb PID system is composed of two ring-imaging Cherenkov detectors (RICH), a series of muon chambers and a calorimeter system (ECAL and HCAL). Combining information from these subdetectors allows one to distinguish between various species of long-lived charged particles. Advanced machine learning techniques are employed to obtain the best PID performance and control systematic uncertainties in a data-driven way. This poster covers the major steps of the implementation, and highlights the PID performance achieved in Run 2.

Global Particle Identification

Particle identification plays a crucial role in high-energy physics analysis. Global PID at LHCb identifies the charged particle type associated with a given track. There are five particle types: electron, muon, pion, kaon, proton, and ghost track. Ghost tracks are charged tracks that do not correspond to a real particle which passed through the detector. Different particle types have different responses in the LHCb systems.

PID is a multiclassification problem in machine learning. Information from the LHCb tracking system, RICHs, calorimeters and muon chambers are used as inputs for the following classifiers to estimate a track type:

- Baseline
- Deep NN: a deep neural network of TMVA library.
- ProdBNN: [1] (baseline) is an one hidden layer neural network of TMV A library;
- [2] is a gradient boosting over oblivious tree model, the Flat 4d model, which is a boosted decision trees classifier, has a flatter PID efficiency as a function of particle $p$, $p_T$, $\eta$ and nTracks (event multiplicity) observables. The classifier achieves this flatness using a modified loss function [3].

The PID information strongly depends on the kinematic variables. This relationship leads to strong dependency between PID efficiency and kinematic variables as shown in Fig 3. Relative to the baseline model, the Flat 4d model, which is a boosted decision trees classifier, has a flatter PID efficiency as a function of particle $p$, $p_T$, $\eta$ and nTracks (event multiplicity) observables. The classifier achieves this flatness using a modified loss function [3].

Conclusions

Combining information from the LHCb tracking system, ring-imaging Cherenkov detectors, electromagnetic and hadron calorimeters, and muon chambers using advanced machine learning techniques allows to achieve high quality of global charged particle identification.

References