Machine Learning approach to γ/π^0 separation in the LHCb calorimeter

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Machine Learning approach to γ /π^0 separation in the LHCb calorimeter

Viktoria Chekalina and Fedor Ratnikov [1]

National Research University Higher School of Economics,
20 Myasnitskaya Str., Moscow, Russian Federation
Yandex School of Data Analysis,
16 Lva Tolstogo Str., Moscow, Russian Federation

E-mail: `vicusya22@gmail.com`

Abstract. Reconstruction and identification of particles in calorimeters of modern High Energy Physics experiments is a complicated task. Solutions are usually driven by a priori knowledge about expected properties of reconstructed objects. Such an approach is also used to distinguish single photons in the electromagnetic calorimeter of the LHCb detector at the LHC from overlapping photons produced from decays of high momentum π^0 . We studied an alternative solution based on first principles. This approach applies neural networks and classifier based on gradient boosting method to primary calorimeter information, that is energies collected in individual cells of the energy cluster. Mutial application of this methods allows to improve separation performance based on Monte Carlo data analysis. Receiver operating characteristic score of classifier increases from 0.81 to 0.95, that means reducing primary photons fake rate by factor of two or more.

1. Introduction

Reconstruction and identification in calorimeters of modern High Energy Physics experiments is a complicated task. The particular problem is to separate prompt photons produced in collisions and photons produced by π^0 which decay into two photons. The most confusing case is when two photons from the π^0 decay have so close directions, that they appear similar to a single cluster produced by prompt photons. Direct solutions are usually driven by a priori knowledge about expected properties of reconstructed objects. Currently used baseline solution of this problem for data collected by the LHCb experiment [1, 2] relies on variables characterising shape of the cluster produced in the electromagnetic calorimeter. In this paper we describe an exercise to separate those two signals without prior knowledge of particular discriminating features of the analysed cluster but rather train algorithm on plain energy deposits in the cluster cells around the cluster seed. This is an attempt to test if modern Machine Learning algorithms may be applied to low level experimental data. We call it *ab initio* approach.

2. Baseline Approach

LHCb Electromagnetic Calorimeter system [1] consists of Scintillating Pad Detector, Preshower Detector (PS), and shashlik-type electromagnetic calorimeter *per se* (ECAL). Energies of PS

---

[1] On behalf of the LHCb Collaboration

and ECAL cells in vicinity of the high energy "seed" ECAL cell are used to discriminate merged π^0 from prompt photons.

It is assumed that the main discriminating feature separating photon and merged π^0 clusters is a cluster shape. Merged π^0 clusters are expected to be elongated and asymmetric due to residual offset between two photons, while genuine photon clusters are expected to be more symmetrical.

Considering a cluster of N cells, with (x_c, y_c) to be coordinates of the cluster center of gravity, e_i to be energy of the ith cell, and (x_i, y_i) to be the cell coordinates. Then the matrix describing second moments of the cluster shape is defined as :

$$S_{xx} = \frac{\sum_{i=1}^{N} e_i(x_i - x_c)^2}{\sum_{i=1}^{N} e_i}, \; S_{yy} = \frac{\sum_{i=1}^{N} e_i(x_i - x_c)^2}{\sum_{i=1}^{N} e_i}, \; S_{xy} = S_{yx} = \frac{\sum_{i=1}^{N} e_i(x_i - x_c)(y_i - y_c)}{\sum_{i=1}^{N} e_i}$$

Features which are used for the baseline separation approach are defined as follows :

- $r_2 = \langle r \rangle = S_{xx} + S_{yy}$
- $r_2, r_4 = 1 - \frac{\langle r^2 \rangle^2}{\langle r^4 \rangle}$
- $k = \sqrt{(1 - 4\frac{S_{xx}S_{yy} - S_{xy}^2}{(S_{xx} + S_{yy})^2}}$

- $asym = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$
- $\frac{E_{seed}}{E_{cl}}$
- $\frac{E_{seed} + E_{snd}}{E_{cl}}$

A 2-layer perceptron was applied to separate photon and π^0 clusters by these features built using energies in 3×3 ECAL cells area around the cluster seed . Full description of the baseline approach may be found in [3].

The training photon sample was done by K*gamma photons and set of π^0 from B decays. The performance of the baseline approach on the reference data described below is presented as dashed line in Fig. 1.

3. Ab initio approach

To separate photons and merged π^0 we consider energies in 5×5 ECAL and PS cells around the cell seed. However we do not build any sophisticated features based on physics considerations, but rather consider plain values of energy allocated in every cell of the 5×5 matrix in both ECAL and PS as a feature. These are 50 plain features to be used by the classifier.

We start with using 2-layer Neural Network which has 2 parallel branches of 2 full-connected layers for preprocessesing information from electromagnetic calorimeter and preshower. For decision tree approach we use XGBoost classifier with 6000 trees. To prevent overfitting, the tree's depth is restricted by 3.

Photons and π^0 used for training are produced in $B^0 \rightarrow K\pi\gamma$ and $B^0 \rightarrow K\pi\pi^0$ processes respectively, and cluster information is extracted from corresponding MC samples. To prove stability of obtained numbers we also use another π^0 sample: $B^0 \rightarrow J\psi K^\star$ with $K^\star \rightarrow K\pi^0$. For this study we only consider high energy clusters with transversal energy $E_T > 2 \; GeV$. We also selected potentially merged π^0 by requiring for true MC photons from π^0 decay to be separated by not more than two calorimeter cells apart on the face of the ECAL.

The procedure of selecting and tuning architecture of the classifier is described in Sec. 4. To find enough number of trained epoch for neural network we plot train-test score curve for every step and noticed when it came out on the plateau. BDT-based classifiers demonstrate better performance for our problem, so we use the XGBoost [4] tool to train and test classifier. Fig. 1 demonstrates performance of the new approach in comparison with the baseline one. The score under Receiver operating characteristic curve improves from 0.89 for the baseline up to 0.97 in new approach. Considering 98% photon efficiency, new approach reduces fake rate from about 60% to about 30%.
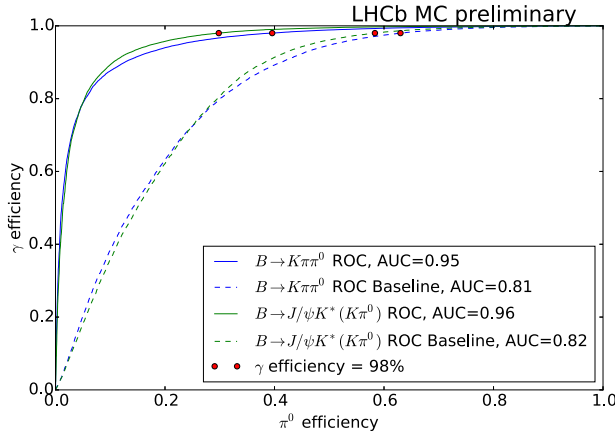
Figure 1. Classificator ROC curves for the baseline (dashed line) and new method (solid line). Different colours refer to different test samples.

Flatness is an important characteristics of the classifier, as it directly affects systematic uncertainties for physics analyses. Fig.2 demonstrates that new approach also have very good flatness in E_T.

3.1. Edge effects

LHCb electromagnetic calorimeter consists of three regions with cell sizes of 4×4, 6×6 and 12×12 cm^2 for inner, middle and outer calorimeter regions respectively. Regions are rectangular, thus the most of the calorimeter area contains the regular grid, and 5×5 clusters can be easily selected. However, every region still has edges which breaks regular structure. For two-sided borders we build complete cluster by upscaling or downscaling energy responses in the adjacent region. Clusters adjacent to innermost and outermost calorimeter remain incomplete. Generally, we have 3 different type of objects: full, reconstructed and incomplete. As properties of diversed cluster's type are expected to be different, a dedicated classifiers for border objects, reconstructed and incomplete, are trained. The full classifier discussed in the previous section, thus consists of 9 individually trained classifiers corresponding to three calorimeter regions, serving regular area, inner and outer edges for each region.
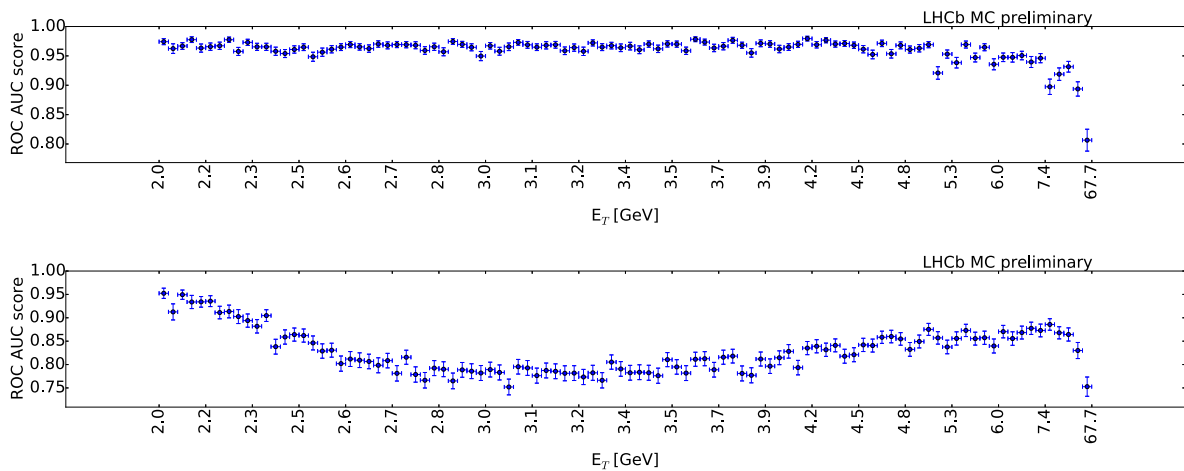

Figure 2. ROC AUC scores in different E_T intervals for baseline (bottom) and new (top) approaches.
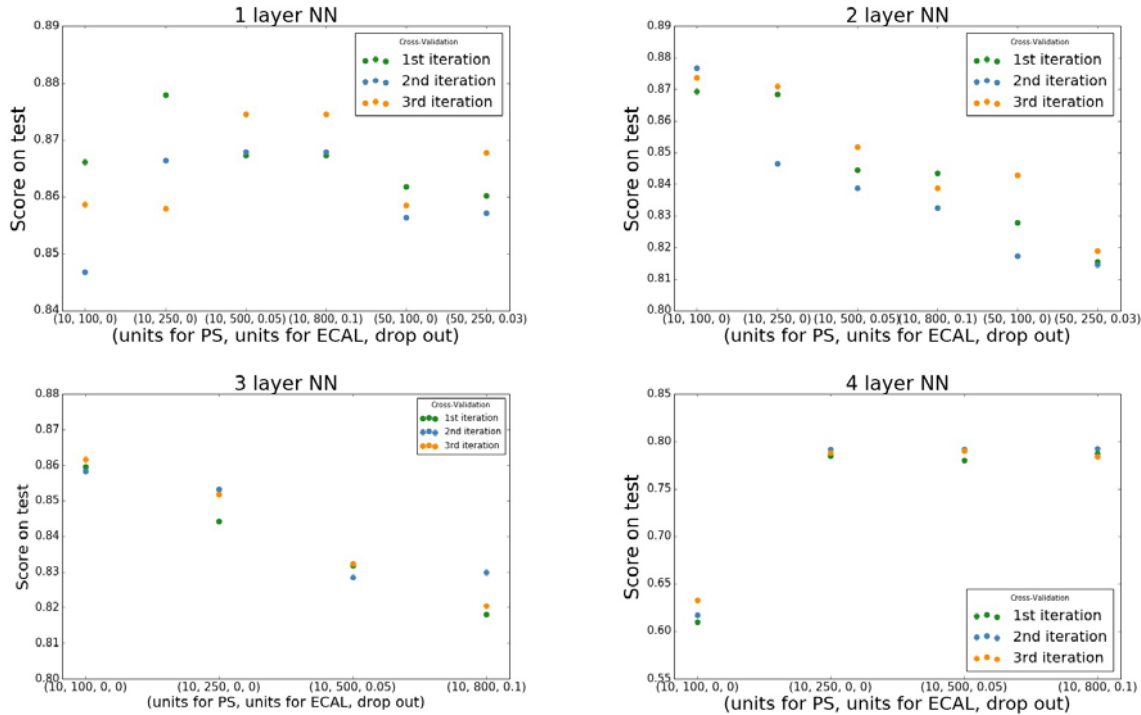
Figure 3. Cross Validation for network configurations with 1, 2, 3, 4 hidden layers. Quality for different iterations is presented as a function of (units for PS branch, units for ECAL branch, dropout).
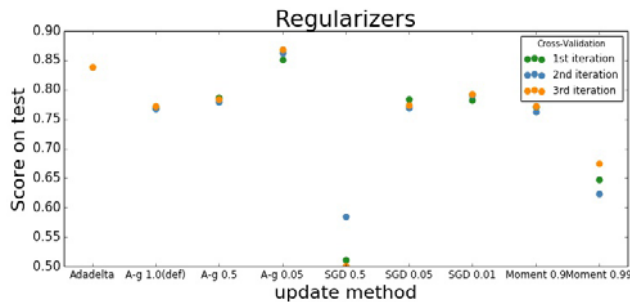

Figure 4. NN classifier quality for various kinds of update functions for 2-layer NN with (units for PS branch, units for ECAL branch, dropout)=(10, 100, 0)

4. Optimisation of the classifier architecture

To select the best classifier architecture we first have to compare Neural Network (NN) and Decision Tree (BDT) approaches. Regular area of the inner calorimeter is used as a reference for quantitive comparison of different classifiers

We start with optimising the possible NN architecture. As behaviours of PS and ECAL detectors are different, and ECAL is stacked after PS concerning the direction of the incoming particles, we build network consisting of two parallel branches of two full-connected layers separately preprocessesing information from ECAL and PS. We found that PS information is not as informative as ECAL. This could easily lead to overtraining. However we also found that using PS branch with smaller number of units could improve quality of the classifier.

When we consider classifiers structure, there are many hyper-parameters which should be optimised to get better quality. For neural networks, besides network topology hyper-parameters include: number of layers and number of units in each layer, method of regularisation, activation
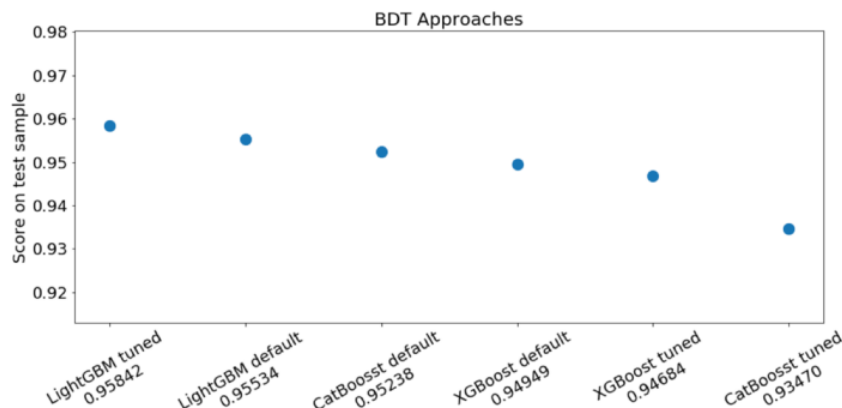
Figure 5. BDT classifiers quality for different BDT implementations with default and tuned set of hyper-parameters.

function, *etc.* Assuming that NN topology described above is fixed, we tune other hyper-parameters to get the best quality. We use 3-fold cross-validation and train network during 500 epoch. The later is about where the plateau for quality improvement is obtained.

Fig. 3 presents cross validation results for different NN configurations with 1-4 hidden layers. Despite of dropout, quality for 3 and 4 layers configuration degrades, that we explain by not too complicated structure of input features.

Results for final variating classifier optimisation method to get the best quality for the given number of steps is presented in Fig. 4.

For the BDT approach, we assayed XGBoost, CatBoost and LightGBM classifiers. For each classifier we found best configuration through tuning boosting parameters by ModelGym [5]. Fig. 5 presents results for both default or tuned configurations. The obtained quality for different classifiers is found to be very similar and significantly exceeds quality obtained with the best NN configuration, so we use default XGBoost (estimators=2000, learning rate=0.05, max depth = 5, min child weight = 2) to build new, ab initio classifier which was used in Sec. 3.

5. Conclusions

We introduced a new, ab initio method for γ/π^0 separation in calorimeter LHCb. It uses calorimeter cell energies in 5×5 area around cluster seed cell. Several neural network configurations and gradient boosting methods are tested, and XGBoost is used to evaluate results. After test on Monte Carlo samples, the method demonstrates an ability to significantly improve separation quality, and has no significant energy dependence.

References
[1] LHCb collaboration, A. A. Alves Jr.*et al.*, *The LHCb detector at the LHC*, JINST **3** (2008) S08005.
[2] LHCb collaboration, R. Aaij *et al.*, *LHCb detector performance*, Int. J. Mod. Phys. **A30** (2015) 1530022, arXiv:1412.6352.
[3] M. Calvo, E. Cogneras, O. Deschamps, M. Hoballah, *A tool for γ π^0 separation at high energies*, LHCb-PUB-2015-016.
[4] XGBoost: A Scalable Tree Boosting System. `arXiv:1603.02754 [cs.LG]`, March 2016
[5] Model Gym, (2017), GitHub repository, https://github.com/yandexdataschool/modelgym