Application of Machine Learning techniques for particle identification in relativistic heavy ion collisions

Arun Kumar Yadav[1,2], Raktim Mukherjee[3], Subhasis Chattopadhyay[1,2], and Partha Pratim Bhaduri[1,2]
[1]*Experimental High Energy Physics & Applications Group, Variable Energy Cyclotron Centre, Kolkata - 700064, INDIA*
[2]*Homi Bhabha National Institute, Mumbai,400064, INDIA and*
[3] *St. Xavier's College (Autonomous), Kolkata**

Introduction

In relativistic heavy-ion collision experiments, the final state hadrons measured by different detectors are used to characterize the hot and dense strongly interacting fireball produced in the initial state of the collisions. One important and non-trivial task in this endeavour is to identify the hardons using dedicated detectors. In the present work, we have used a simulated data set for ALICE detector system at LHC to compare the performance of the traditional method of particle identification with that from the machine learning models. We use three different machine learning models: a multilayer perceptron (MLP), an adaptive boosting decision tree (BDT), and a gradient boosting decision tree(BDTG).

In the traditional method, protons and pions can be separated based on their specific energy loss in the Time Projection Chamber (TPC) detector, but the response curves for protons and pions in TPC get almost inseparable for particle momentum above 2 GeV/c. Additionally, requiring information from the Time Of Flight(TOF) along with the TPC can overcome this difficulty to some extent. This is traditionally called the combined number of sigma method, where a selection cut is applied combining information from TPC + TOF to discriminate between protons and pions. The selection cut can be made tighter to get a purer sample of target particle(say proton) to be identified, but with a lower identification efficiency. This is because the imposition of a tighter selection cut would reject a lot of actual protons as well. The goal of the particle discrimination technique is to optimize this trade-off between identification efficiency and purity. Details of the ALICE detector system and the traditional methods for particle identification can be found in [1]

Traditional cuts can only be used for univariate analysis and thus often make the extraction of data with high efficiency and purity difficult. Machine Learning models use multivariate analysis techniques and apply a non-linear decision boundary, which could improve efficiency and purity. It is worth investigating if the available Machine Learning tools can provide a better alternative to the combined number of sigmas method by improving upon the purity and particle identification efficiency. We have used the following machine learning tools:

Multilayer perceptron (MLP) is a type of artificial neural network that is commonly used for classification tasks. It is a powerful model that can learn complex relationships between the input and output variables. However, it can be computationally expensive for training and may not be suitable for large data sets.

Adaptive boosting decision tree (AdaBoost) is a machine learning algorithm that combines multiple weak learners to create a strong learner. The weak learners are typically decision trees. AdaBoost is known for its ability to improve the performance of weak

*Electronic address: `ak.yadav@vecc.gov.in`

learners. However, it can be prone to over fitting.

Gradient boosting decision tree (GBDT) is another machine learning algorithm that combines multiple weak learners to create a strong learner. The weak learners are also typically decision trees. GBDT is similar to AdaBoost, but it uses a different approach to combine the weak learners. GBDT is known for its ability to generalize well to unseen data.

The implementation of these models has been done using ROOT's TMVA library[2]. The library has been specifically created for use in High Energy Physics data analysis.

We have applied these models to discriminate between protons and pions, with protons as the particle of interest. ALICE Monte Carlo for Pb+Pb 5.02 TeV collisions have been used which uses HIJING event generator. We have selected the events belonging to the centrality class 0-10%. The generated particles are passed through the simulated detector set up which is then used to train the ML models. The target particle is then identified using the appropriate selection cut for each model. The ratio of primary protons identified after the selection cut to the number of primary protons generated gives the identification efficiency of the model. The purity of the identified sample is he fraction of true protons within the sample of all the particles that satisfy the selection cut. We plot the purity and efficiency as a function of transverse momentum(p_T).

Results and Discussion

The particle identification efficiency of different methods used is presented in fig. 1. We observe the particle identification efficiency to be the best for the BDT method over the entire range of p_T. We also observe that the efficiency corresponding to the traditional combined number of sigma method is lower than all of the ML models used. The purity of protons as obtained using different models is presented in fig .2. The traditional cut-based method gives higher purity but with with a lower efficiency. The centrality dependence of efficiency and purity with higher statistics data will be presented at the symposium.

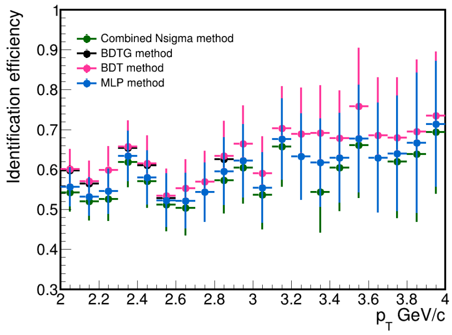

FIG. 1: Identification efficiency as a function of transverse momentum for various methods.

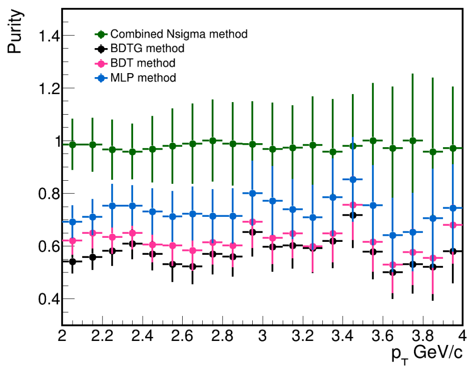

FIG. 2: purity as a function of transverse momentum for various methods.

References

[1] D. Sarkar [CERN-THESIS-2021-317]

[2] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, “TMVA: Toolkit for Multivariate Data Analysis,” PoS A CAT 040 (2007) [physics/0703039].