Estimation of impact parameter and transverse spherocity in heavy-ion collisions at the LHC using machine learning techniques

Neelkamal Mallick[1,*], Sushanta Tripathy[3],
Aditya Nath Mishra[4], Suman Deb[1], and Raghunath Sahoo[1,2]
[1] *Department of Physics, Indian Institute of Technology Indore, Simrol, Indore 453552, India*
[2] *CERN, CH 1211, Geneva 23, Switzerland*
[3] *INFN - sezione di Bologna, via Irnerio 46, 40126 Bologna BO, Italy and*
[4] *Wigner Research Center for Physics, H-1121 Budapest, Hungary*

1. Introduction

The studies related to heavy-ion collisions at the Large Hadron Collider (LHC) at CERN, Switzerland and Relativistic Heavy Ion Collider (RHIC) at BNL, USA have revealed the formation of a dense and hot, deconfined state of matter known as the quark-gluon plasma (QGP). Presently, machine learning (ML) techniques are being used widely in the field of high energy physics (HEP) as well as other frontiers of science. The development of smart algorithms gives machine the ability to learn from training data and to predict outcomes on independent data without being explicitly programmed to do so. It can learn the hidden patterns based on the correlations between the input and the output variables.

In heavy-ion collisions, the impact parameter is such a crucial observable which directly affects the final state particle production. However, it is almost impossible to estimate the impact parameter in such collisions as the length scale associated is of the order of femtometers (10^{-15} m). An event shape observable, the transverse spherocity (S_0) has been used in *pp* collisions at the LHC to study the system dynamics and particle production with respect to jetty and isotropic events [1]. In one of our publications [2], we have implemented transverse spherocity in Pb-Pb collisions to study azimuthal anisotropy and found that S_0 is anti-correlated with v_2. Thus, it is important to exploit the ML tools in heavy-ion collisions at LHC energies to estimate these two observables. In this work, we have used final state observables such as the average charged particle multiplicity ($\langle dN_{\rm ch}/d\eta \rangle$), multiplicity in the transverse plane ($\langle N_{\rm ch}^{\rm TS} \rangle$) and average transverse momentum ($\langle p_{\rm T} \rangle$) as input variables to the gradient boosted decision trees (GBDT) algorithm in the ML framework [3, 4] and performed regression on impact parameter and transverse spherocity in Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV in AMPT model [5, 6].

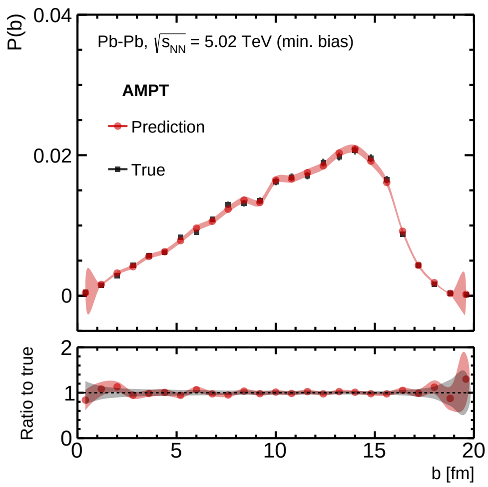

FIG. 1: (Color Online) Predictions for impact parameter distribution using gradient boosted decision trees (GBDT) algorithm in Pb-Pb collisions at $\sqrt{s_{\rm NN}}$ = 5.02 TeV in AMPT model. The bottom panel shows the ratio of true value to model predictions. (Fig. 6 [6])

*Corresponding author: Neelkamal.Mallick@cern.ch

2. Results and Discussions

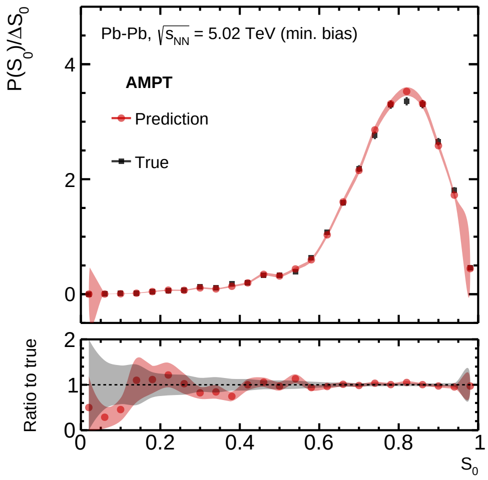

FIG. 2: (Color Online) Predictions for transverse spherocity distribution using gradient boosted decision trees (GBDT) algorithm in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV in AMPT model. The bottom panel shows the ratio of true value to model predictions. (Fig. 7 [6])

The GBDT algorithm offers several tunable parameters which need to be chosen carefully to make the best predictions. Some of these parameters are the number of trees, maximum depth and the learning rate. The task is to obtain a best fit to the training data without compromising the testing accuracy. The accuracy of the model could be evaluated from the mean absolute error of the output variable in the training data. The mean absolute error can be written as,

$$\Delta b = \frac{1}{N_{\rm events}} \sum_{n=1}^{N_{\rm events}} |b_n^{true} - b_n^{pred.}|. \qquad (1)$$

Here, b denotes the impact paramter. Figure 1 shows the impact parameter distribution for Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV in AMPT model. The black points represent the true values from simulation, whereas the red points are the GBDT-ML predictions. Black band shows statistical uncertainties in the true value and red band shows the quadratic sum of statistical and systematic uncertainties in the predicted values. We can see, there is a good agreement between the true and predicted values. Figure 2 represents the predictions for transverse spherocity distribution in Pb-Pb collisions at $\sqrt{s_{\rm NN}} = 5.02$ TeV in AMPT model. It is evident from the ratio plot in the bottom panel that there is a good agreement between the true spherocity values from AMPT model and GBDT-ML predictions.

3. Summary

We report the implementation of gradient boosted decision trees (GBDT) algorithm in machine learning framework for heavy-ion collisions at the LHC using AMPT model. This method is used to predict the impact parameter and a fairly new observable, the transverse spherocity. In this approach, final state observables such as $\langle dN_{\rm ch}/d\eta\rangle$, $\langle N_{\rm ch}^{\rm TS}\rangle$ and $\langle p_{\rm T}\rangle$ have been taken as model input to perform regression on impact parameter and S_0. The method's prediction agrees well with the true simulated values from AMPT model. Model trained at higher energy is capable of estimating S_0 distribution at lower energy for the same collision system. We also found that, the model trained with minimum bias data successfully predicts the individual centrality wise S_0 distributions.

References

[1] S. Acharya *et al.* [ALICE], Eur. Phys. J. C **79** (2019) no.10, 857
[2] N. Mallick, R. Sahoo, S. Tripathy and A. Ortiz, J. Phys. G **48** (2021), 045104
[3] J. H. Friedman, Annals Statist. **29** (2001) no.5, 1189-1232
[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, G. Louppe, P. Prettenhofer and R. Weiss, *et al.* J. Machine Learning Res. **12** (2011), 2825-2830
[5] Z. W. Lin, C. M. Ko, B. A. Li, B. Zhang and S. Pal, Phys. Rev. C **72** (2005), 064901
[6] N. Mallick, S. Tripathy, A. N. Mishra, S. Deb and R. Sahoo, Phys. Rev. D **103** (2021), 094031 and references therein.