Classifier for centrality determination with Zero Degree Calorimeter at the Cooling-Storage-Ring External-target Experiment

Authors: Biao Zhang, Li-Ke Liu, Hua Pei, Shusu Shi, Nu Xu, Yaping Wang, Yaping Wang

Date: 2023-10-24T00:00:00+00:00

Abstract

The Zero Degree Calorimeter (ZDC) plays a crucial role in determining centrality at the Cooling-Storage-Ring External-target Experiment (CEE) in the Heavy Ion Research Facility in Lanzhou (HIRFL). A Boosted Decision Trees (BDT) multi-classification algorithm is employed to classify the centrality of the collision events based on the raw features from ZDC such as the number of fired channels and deposited energy. The data from simulated \$ rm ^{238}U\$ + \$ rm ^{238}U\$ collisions at 500 \$ rm MeV/u\$, generated by the IQMD event generator and subsequently modeled through the GEANT4 package, is employed to train and test the BDT model. The results showed the high accuracy of the multi-classification model adopted in ZDC for centrality determination, which is robust against variations in different factors of detector geometry and response. The study demonstrates a good performance of the CEE-ZDC for determining the centrality in nucleus-nucleus collisions.

Full Text

Preamble

Classifier for Centrality Determination with Zero Degree Calorimeter at the Cooling-Storage-Ring External-target Experiment

Biao Zhang, Li-Ke Liu, Hua Pei, Shu-Su Shi, Nu Xu, and Ya-Ping Wang

Key Laboratory of Quark & Lepton Physics (MOE) and Institute of Particle Physics, Central China Normal University, Wuhan 430079, China

The zero-degree calorimeter (ZDC) plays a crucial role in determining centrality in the Cooling-Storage-Ring External-target Experiment (CEE) at the Heavy Ion Research Facility in Lanzhou (HIRFL). A boosted decision tree (BDT)

multi-classification algorithm was employed to classify the centrality of collision events based on raw features from the ZDC, such as the number of fired channels and deposited energy. Data from simulated ^{238}U + ^{238}U collisions at 500 MeV/u, generated by the IQMD event generator and subsequently modeled using the GEANT4 package, were used to train and test the BDT model. The results demonstrated high accuracy of the multiclassification model for centrality determination with ZDC, which remains robust against variations in detector geometry and response factors. This study demonstrates the excellent performance of CEE-ZDC in determining centrality in nucleus-nucleus collisions.

Keywords: ZDC, Boosted Decision Trees, Multi-classification, IQMD, Centrality determination

Introduction

The primary objective of heavy-ion collisions at different beam energies is to investigate strongly interacting matter and understand the QCD phase diagram. This diagram provides information about phase transitions and critical points in strongly interacting systems, where hadron gases exist at lower temperatures and baryon densities. At higher temperatures or densities, the hadronic boundary disappears and confined quarks move freely throughout the system [**?**]. The Beam Energy Scan program of RHIC-STAR aims to approach the possible critical point from the high-energy side, while it is essential to study the hadronic phase diagram and approach the critical point from the low-energy side [**?**, **?**, **?**]. The Cooling-Storage-Ring External-target Experiment (CEE) at the Heavy Ion Research Facility in Lanzhou (HIRFL), with its advanced spectrometer, provides significant opportunities for studying phase diagrams at extremely high net baryon density levels with beam energies of several hundred AMeV [**?**].

The zero-degree calorimeter (ZDC), one of the CEE subdetectors in the forward rapidity region, is designed to accurately determine the centrality and reaction plane of collision events [**?**]. Collision events are typically classified into centrality classes representing certain fractions of the total reaction cross-section corresponding to specific intervals of the impact parameter b [**?**]. The impact parameter b is essential for understanding the initial overlap region of colliding nuclei in heavy-ion collisions; it represents the distance between the nuclear centers in the plane transverse to the beam axis and determines the size and shape of the resulting medium. However, the impact parameter b is not directly measurable in experiments. To estimate centrality experimentally, raw observables that scale monotonically with impact parameter can be used for classification, such as reconstructed tracks from central barrel tracking detectors or deposited energy in forward calorimeters. Accurate centrality determination is a baseline requirement for many physical analyses in heavy-ion collision experiments [**?**], particularly when searching for observables sensitive to possible phase transitions or critical points through fluctuation and correlation analyses.

In recent years, machine learning (ML) methods have gained significant attention for determining the centrality of heavy-ion collisions [?, ?]. Previous studies treated centrality determination as a regression problem on impact parameters and utilized combined information from central tracking systems and forward calorimeters to train ML models. However, to avoid autocorrelation in physics analyses, this study adopts a machine-learning approach that utilizes only raw experimental features from a forward calorimeter to determine centrality. We report the application of a multi-classification ML algorithm based on Boosted Decision Trees (BDT) as a centrality classifier using only ZDC data from ^{238}U + ^{238}U collisions at 500 MeV/u at CEE. The ML inputs were generated using the Isospin-dependent Quantum Molecular Dynamics (IQMD) generator [?], and we present the efficiency and purity measures related to the centrality determination performance of the ZDC with this model application.

CEE-ZDC

CEE utilizes fixed-target-mode heavy-ion collisions and is the first large-scale experimental nuclear physics facility operating in the GeV energy region in China. It is equipped with a comprehensive set of subdetectors, as shown in Fig. 1 Figure 1: see original paper. The detector system comprises a beam monitor, T0 detector [?], time projection chamber (TPC) [?], inner time-of-flight (iTOF) detector [?], large superconducting dipole magnet, multiwire drift chamber (MWDC) [?], external time-of-flight (eTOF) detector [?], and zero-degree calorimeter (ZDC) [?]. The ZDC is designed to detect particle fragments in the forward rapidity region following semi-central and peripheral collisions, providing vital information for precise reconstruction of centrality and reaction plane [?, ?]. Positioned at the downstream end of CEE, the ZDC covers a pseudorapidity range of $1.8 < \eta < 4.8$ and utilizes a symmetrical, fan-shaped layout with eight radial and 24 angular sections extending to a maximum radius of 1 m. The detector comprises trapezoidal modules equipped with uniform plastic scintillators coupled to light guides and photomultiplier tubes (PMTs) that convert scintillation light into charge signals. To obtain comprehensive signal information, each module provides two charge signals from the two dynodes of each PMT that are transmitted to separate readout channels, resulting in 384 ($24 \times 8 \times 2$) channels for the ZDC.

Model Training with Simulated Events

The simulated data were generated by modeling ^{238}U + ^{238}U collisions at 500 MeV/u using the IQMD generator [?], with subsequent particle transport through the apparatus simulated using the GEANT4 package [?]. Determining centrality with only one forward rapidity detector such as ZDC is challenging even when employing ML algorithms. Previous ML-based studies on centrality determination relied on information from multiple detector subsystems, such as tracks reconstructed from central barrel detectors and deposited energy in forward calorimeters, revealing strong correlations between centrality class and observables. CEE-ZDC is a non-tracking detector, and the number of spectator nucleons in a nucleus-nucleus collision is expected to be proportional to the deposited energy and number of fired channels in the ZDC. However, the presence of a beam hole at the center of ZDC and limited detector acceptance result in weak monotonic dependence between impact parameter and observables, as illustrated in Fig. 2 Figure 2: see original paper for the number of fired channels and Fig. 2(b) for deposited energy in ZDC.

Potential improvements in centrality determination can be achieved by utilizing data from ZDC subrings as an additional feature in the ML task. Moreover, it may be advantageous to use the energy deposited in the ZDC ring-by-ring and the number of event-by-event fired channels, exploiting all inherent correlations between modules. Fig. 3 Figure 3: see original paper displays the probability distribution of fired ZDC channels in the impact parameter range $7 < b \leq 10$ fm, while Fig. 3(b) shows the probability distribution of deposited energy in ZDC rings for $0 < b \leq 3$ fm. The complex patterns and nontrivial decision boundaries among event centrality classes present an ideal opportunity for applying ML techniques.

Boosted Decision Trees (BDT) represent a family of popular supervised learning algorithms for classification and regression problems that are extensively used in high-energy physics data analysis. In this study, extreme gradient boosting (XGBoost), a powerful BDT implementation based on the gradient boosting method, was adopted to solve the multi-classification problem for centrality determination. The physical features used as inputs for model training include the deposited energy in the full ZDC and ZDC subrings, as well as the number of fired channels in ZDC. The simulated data were divided into three centrality classes based on impact parameters listed in Table 1 . Samples were split into equal-sized training and test sets for each centrality class. State-of-the-art hyperparameter optimization with Optuna was employed to accelerate optimization and achieve optimal model performance [**?**].

Performance of the ML Models

The machine learning model was applied to both training and test sets to visualize distributions of ML output scores and verify consistency between the two sets. For classification with three centrality classes (p_i), the model generates three scores representing the probability of belonging to each class. By construction, the probabilities for the centrality classes sum to one ($\sum_{i=1}^{3} p_i = 1$). Fig. 4 [Figure 4: see original paper] illustrates the probability distributions for the central (a) and peripheral (b) classes for both training and test sets. The probability distributions peak near unity for the respective true class, while the other two distributions shift toward zero. The probability density functions of training and test samples for each centrality class agree well, indicating that the model does not overfit.

The Receiver Operating Characteristic (ROC) curve is commonly used to evaluate classification model performance by plotting the true-positive rate against the false-positive rate for various threshold settings. The area under the ROC curve, known as ROC AUC, provides a global measure of model performance ranging from 0.5 (random classification) to 1 (perfect classification), independent of threshold and class distribution [?]. For multi-class classification, the ROC curve cannot be directly defined, and the "One-vs-One" approach is used to compute the overall average of individual ROC AUCs for each pair of classes. The ROC curves and ROC AUC values obtained for the test set are shown in Fig. 5 [Figure 5: see original paper]. The high final ROC AUC value of approximately 0.96 indicates that the BDT model is highly effective for centrality determination.

Efficiency and Purity of the Centrality Classification

The performance of the centrality classification model was evaluated by calculating its efficiency and purity based on ML output scores. Efficiency refers to the fraction of correctly classified events, whereas purity measures the fraction of events correctly classified for a particular centrality class out of all events assigned to that class. The efficiency versus purity curves for the multi-classification models for each centrality class are shown in Fig. 6 [Figure 6: see original paper], where red, green, and blue solid lines represent central, semi-central, and peripheral classes, respectively. The peripheral class was most effectively classified, while the central class proved more challenging than the semi-central class in higher-efficiency regions. The values listed in Table 2 indicate that even at very high purity levels, the efficiency of the peripheral class is not significantly compromised, and both central and semi-central classes exhibit promising efficiency values at high purity. These results demonstrate that ML-based event centrality determination utilizing ZDC is effective.

To evaluate the robustness of centrality determination with ZDC, the effects of several factors related to ZDC configuration in simulation data were systematically investigated. These factors include the thickness of the ZDC plastic scintillator, hit efficiency, energy resolution, and heavy nuclei with or without de-excitation (tunable settings in IQMD). The ZDC plastic scintillator thickness was varied from 1 to 4 cm, and hit efficiency was varied from 90% to 95%. The deposited energy was also smeared with different sigma values of Gaussian distributions. As illustrated in Fig. 7 [Figure 7: see original paper], red, green, and blue lines indicate central, semi-central, and peripheral collisions, respectively, with changes in these factors depicted by distinct line styles. The results indicate that these factors have minor effects on the purity and efficiency of centrality classification. Among the tested factors, ZDC detector thickness had the most significant impact, though this effect was relatively small. This study suggests that the multi-classification approach adopted for ZDC is robust against variations in these factors, indicating potential for reliable and accurate centrality classification using ZDC.

Summary

This study aimed to determine the centrality class of nucleus-nucleus collisions at the CEE-ZDC detector using a multi-classification model based on the XGBoost classifier. The ML model was trained and tested using simulation data from the IQMD event generator and subsequently modeled using the GEANT4 package. An additional study examined various factors associated with ZDC detector geometry and response. The results indicated that these factors had minor impact, demonstrating the robustness of the XGBoost classifier for centrality determination. Future work may include improving centrality determination accuracy by incorporating regression tasks and exploring other machine-learning algorithms. This study demonstrates the good performance of CEE-ZDC for centrality determination in nucleus-nucleus collisions.

Acknowledgments

We thank Prof. Li Ou and Zhigang Xiao for generating IQMD data and for fruitful discussions.

References

[1] P.B. Munzinger and J. Stachel, The quest for the quark-gluon plasma. Nature 448, 302–309 (2007). doi: 10.1038/nature06080
[2] D. Oliinychenko, A. Sorensen, V. Koch et al., Sensitivity of Au+Au collisions to the symmetric nuclear matter equation of state at 2 − 5 nuclear saturation densities. arXiv:2208.11996
[3] H.Z. Huang, F. Liu, X.F. Luo et al., Collective excitation in high-energy nuclear collisions – in memory of Professor Lianshou Liu. Symmetry 15, 499 (2023). doi: 10.3390/sym15020499
[4] X. Luo, S. Shi, N. Xu et al., A study of the properties of the QCD phase diagram in high-energy nuclear collisions. Particles 3, 278–307 (2020). doi: 10.3390/particles3020022
[5] D.D. Hu, J.M. Lu, J. Zhou et al., Extensive beam test study of prototype MRPCs for the T0 detector at the CSR external-target experiment. Eur. Phys. J. C 80, 282 (2020). doi: 10.1140/epjc/s10052-020-7804-2
[6] S.H. Zhu, H.B. Yang, H. Pei et al., Prototype design of readout electronics for Zero Degree Calorimeter in the HIRFL-CSR external-target experiment. JINST 16, P08014 (2021). doi: 10.1088/1748-0221/16/08/P08014
[7] A. Svetlichnyi, R. Nepeyvoda, I. Pshenichnov, Using Spectator Matter for Centrality Determination in Nucleus-Nucleus Collisions. JINST 4, 227–235 (2021). doi: 10.3390/particles4020021
[8] A. Igor and K. Vladimir, Classifiers for centrality determination in proton-nucleus and nucleus-nucleus collisions. EPJ Web Conf. 137, 11001 (2017). doi: 10.1051/epjconf/201713711001
[9] M. Omana Kuttan, J. Steinheimer, K. Zhou, A. Redelbach, H. Stoecker, A fast centrality-meter for heavy-ion collisions at the CBM experiment. Phys.

Lett. B. 811, 135872 (2020). doi: 10.1016/j.physletb.2020.135872
[10] C. Hartnack, R.K. Puri, J. Aichelin et al., Modeling the many body dynamics of heavy ion collisions: Present status and future perspective. Eur. Phys. J. A 1, 151-169 (1998). doi: 10.1007/s100500050045
[11] W. Huang, F. Lu, H. Li et al., Laser test of the prototype of CEE time projection chamber. Nucl. Sci. Tech. 29, 41 (2018). doi: 10.1007/s41365-018-0382-4
[12] X. Wang, D. Hu, M. Shao et al., CEE inner TOF prototype design and preliminary test results. JINST 17, P09023 (2022). doi: 10.1088/1748-0221/17/09/P09023
[13] L. Lyu, H. Yi, L. Duan et al., Simulation and prototype testing of multi-wire drift chamber arrays for the CEE. Nucl. Sci. Tech. 31, 11 (2020). doi: 10.1007/s41365-019-0716-x
[14] B. Wang, D. Han, Y. Wang et al., The CEE-eTOF wall constructed with new sealed MRPC. JINST 15, C08022 (2020). doi: 10.1088/1748-0221/15/08/C08022
[15] L.-K. Liu, H. Pei, Y.P. Wang et al., Event plane determination from zero-degree calorimeter at the cooling-storage-ring external-target experiment. Nucl. Sci. Tech. 34, 100 (2023). doi: 10.1007/s41365-023-01262-8
[16] R. Brun, F. Bruyant, F. Carminati et al., GEANT detector description and simulation tool (1994). CERN. doi: 10.17181/CERN.MUHF.DMJ1
[17] T. Akiba, S. Sano, T. Yanase et al., Optuna: A Next-generation Hyperparameter Optimization Framework. arXiv: 1907.10902
[18] D.M.W. Powers, Optuna: A Next-generation Hyperparameter Optimization Framework. arXiv: 2010.16061

Note: Figure translations are in progress. See original paper for figures.