Application of machine learning techniques at BESIII experiment

Beijiang Liu[*][†]

Institute of High Energy Physics, Chinese Academy of Sciences
E-mail: liubj@ihep.ac.cn

Xian Xiong

Institute of High Energy Physics, Chinese Academy of Sciences
*E-mail:*xiongxa@ihep.ac.cn

Guoyi Hou

Institute of High Energy Physics, Chinese Academy of Sciences

Shiming Song

Sichuan University

Lin Shen

Sichuan University

BESIII is a currently running tau-charm factory with the largest samples of on threshold charm meson pairs, directly produced charmonia and some other unique datasets at BEPCII collider. Machine learning techniques have been employed to improve the performance of BESIII software. The studies for reweighing MC, particle identification and cluster reconstruction for the CGEM (Cylindrical Gas Electron Multiplier) inner tracker are presented.

---

[*]Speaker.

1. Introduction

The BESIII experiment (Beijing Spectrometer) [2], located at BEPCII (Beijing Electron Position Collider), is an experiment at the high precision frontier of hadron physics in τ-charm region. BESIII has been successfully operating since 2008 and has collected the world's largest data samples of J/ψ, $\psi(3686)$ and $\psi(3770)$ decays as well as data in the energy region above 4 GeV, which benefit rich physics programs [1]. Machine learning (ML) techniques have been employed to improve the performance of BESIII software. Novel approaches for muon identification, multi-dimensional reweighting and cluster reconstruction for the Cylindrical Gas Electron Multiplier inner tracker (CGEM-IT) are presented.

2. A new approach for muon identification

The BESIII detector has a geometrical acceptance of 93% of 4π and consists a small-celled, helium-based main draft chamber, an electromagnetic calorimeter (EMC), a time-of-flight system (TOF) for particle identification, a muon chamber system (MUC) with resistive plate chambers incorpared in the return iron of the superconducting solenoid with 1T magentic field. More details of the detector are described in Ref [2]. A combined confidence level for muon identification is formed by using the specific ionization energy loss in the MDC (dE/dx), the time of flight, and the information from EMC and MUC. $\mathscr{L} = \mathscr{L}_{\mathrm{dE/dx}} \cdot \mathscr{L}_{\mathrm{TOF}} \cdot \mathscr{L}_{\mathrm{EMC}} \cdot \mathscr{L}_{\mathrm{MUC}}$, where $\mathscr{L}_{\mathrm{dE/dx(TOF)}}$ is calculated from χ^2 of particle hypothesis (e, μ, π, K, p) and $\mathscr{L}_{\mathrm{EMC(MUC)}}$ is a normalized output of a shallow neural network of EMC(MUC). The discriminate of μ / π is critical for many of the analyses. The task is very challenging, because the mass of μ and π are very close and μ and π are difficult to be discriminated with dE/dx and TOF.

We propose a nesting architecture with XGBoost [3] classifiers for μ identification. We use two classifiers with all the reconstructed information from EMC and MUC as inputs, respectively. The outputs of the two classifiers together with $\chi^2_{\mathrm{dE/dX}}$ and χ^2_{TOF} are submitted to another classifier for combination. Based on studies with MC samples, Figure 1 shows the new approach with ML has a better performance than the default muon identification used in BESIII. The performance drops around 0.4 GeV because there is a cut-off of the MUC for those low momentum particles cannot reach the muon counter.
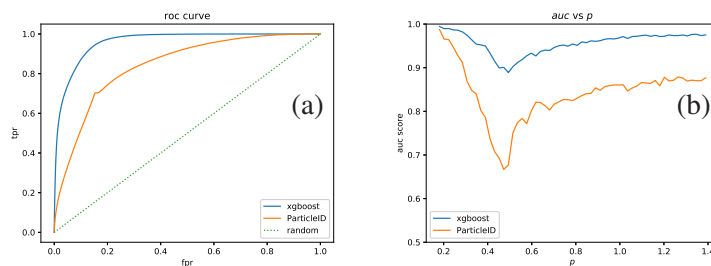

Figure 1: Comparison between new approach with XGBoost (blue curve) and the default PID (yellow curve) (a) ROC curve (b) AUC curve.

3. Reweighting with XGBoost

Modeling data with Monte Carlo (MC) simulation is essential for physics analysis, e.g., for calculation of efficiency and estimation of background. In the energy regime of non-perturbative QCD, experimental data of BESIII is rich of resonances, which cannot be fully described with a generic MC model. Amplitude analysis [4] is usually required to extract the properties of resonances and model the data. In the case that results of amplitude analysis are not available yet, it is very useful to create "data-like" MC by multi-dimensional reweighting to correct for differences between data and MC. There are two methods for reweighting with ML techniques [5, 6] have been proposed. We utilize the approach [5] with XGBoost algorithm for physics analysis at BESIII [7]. For illustration, a MC sample of $J/\psi \to N^*\bar{n} + c.c. \to p\pi^-\bar{n} + c.c.$ including a set of intermediate $p\pi$ resonances is generated as pseudo data. Reweighting is applied to a sample of phase-space-distributed MC (PHSP). Figure 2 shows the results after reweighting.
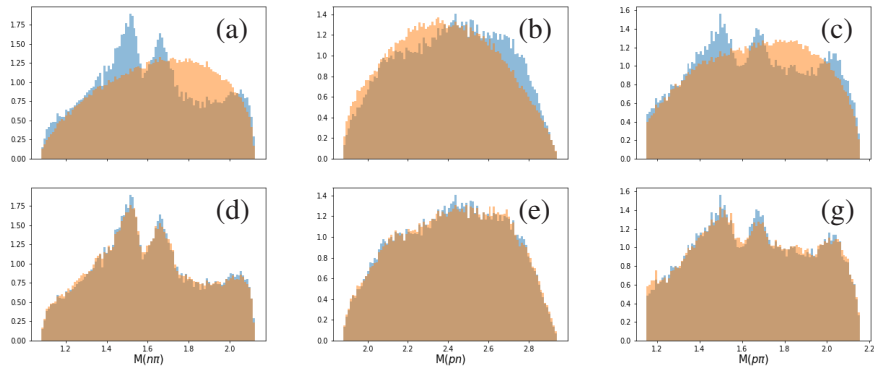

Figure 2: Comparison of pseudo data (blue) and PHSP (yellow) distributions before(a-c) and after (d-f) using the XGBoost reweighter.

4. Cluster reconstruction of cylindrical GEM inner tracker

Due to the aging of the inner drift chamber, an upgrade of inner tracker with 3 layers of cylindrical triple-GEMs is scheduled for BESIII in 2019. As the input of track reconstruction, cluster reconstruction is to measure the position of the ionizing particle in the drift cathode layer with the readouts from the anode strips. There are two methods for cluster reconstruction of CGEM-IT [8]. The charge centroid method calculates the weighted average position of the strips with their charge (Q). The method based on time measurements(T) is using the drift gap as a "micro time projection chamber" (micro-TPC) [9]. To improve the position resolution, the results of the two methods can be further combined according to their resolutions and correlations. However, the correlations between resolution and incident angle are quite complicated and difficult to handle. A ML method based on XGBoost regressor is proposed to reconstruct the initial ionizing particle position from the readouts (Q and T) of the fired strips. A simulation with a standalone digitization code, based on GARFIELD [10], is used to generate the event with 1 T magnetic field, incident angle between -30° to 30° for one layer of planar Triple-GEM. The results compared with the charge centroid method are shown in Figure 3. The resolution of the XGBoost regressor with Q or T inputs alone

is better than that of the charge centroid method. The results with Q have larger dependency with incident angle. The combined result with Q and T by ML is further improved.
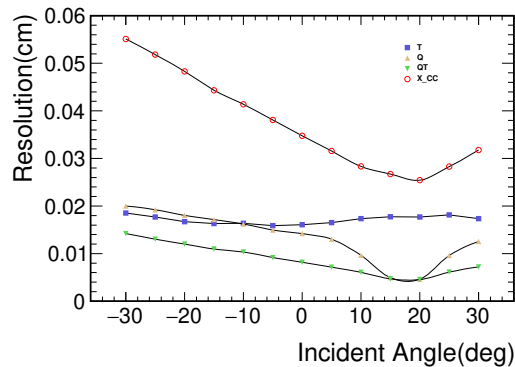

Figure 3: Resolution curve along with incident angle, blue curve for T input only, yellow curve for Q input only, green curve for Q, T input together, red curve for charge centroid results.

5. Summary

Three new use cases of ML at BESIII are presented in this proceeding and show promising results. In the future, we can further investigate the application of ML to improve the performance of BESIII software, e.g., the tracking of low momentum charged particles, the tracking with high background rates, *etc.*

References

[1] D. M. Asner *et al.*, Int. J. Mod. Phys. A **24**, S1-794 (2009)

[2] M. Ablikim *et al.* (BESIII Collaboration), Nucl. Instrum. Methods Phys. Res., Sect. A **614**, 345 (2010)

[3] T. Chen , C. Guestrin, Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (785-794) (2016)

[4] M. Ablikim *et al.* (BESIII Collaboration), Phys. Rev. D **88** 112007, (2013)

[5] D. Martschei *et al.*, Journal of Physics: Conference Series, 368(1), 012028 (2012)

[6] A. Rogozhnikov, Journal of Physics: Conference Series, 762(1), 012036 (2016)

[7] Xian Xiong, & Beijiang Liu. (2018, October 9). Zenodo. http://doi.org/10.5281/zenodo.1451985

[8] R. Farinelli *et al.*, arXiv:1807.00500 (2018)

[9] T. Alexopoulos *et al.*, Nucl. Instrum. Meth. A **617**, 161 (2010).

[10] R.Farinelli, L.Lavezzi *et al.*, arXiv:1807.01210 (2018)