Contribution ID: **5**

Type: **Poster**

Towards more precise data analysis with Machine-Learning-based particle identification with missing data

Identifying products of ultrarelativistic collisions delivered by the LHC and RHIC colliders is one of the crucial objectives of experiments such as ALICE and STAR, which are specifically designed for this task. They allow for a precise Particle Identification (PID) over a broad momentum range.

Traditionally, PID methods rely on hand-crafted selections, which compare the recorded signal of a given particle to the expected value for a given particle species (i.e., for the Time Projection Chamber detector, the number of standard deviations in the dE/dx distribution, so-called "nσ" method). To improve the performance, novel approaches use Machine Learning models that learn the proper assignment in a classification task.

However, because of the various detection techniques used by different subdetectors (energy loss, time-of-flight, Cherenkov radiation, etc.), as well as the limited detector efficiency and acceptance, particles do not always yield signals in all subdetectors. This results in experimental data which include "missing values". Out-of-the-box ML solutions cannot be trained with such examples without either modifying the training dataset or re-designing the model architecture. Standard approaches to this problem used, i.e., in image processing involve value imputation or deletion, which may alter the experimental data sample.

In the presented work, we propose a novel and advanced method for PID that addresses the problem of missing data and can be trained with all of the available data examples, including incomplete ones, without any assumptions about their values [1,2]. The solution is based on components used in Natural Language Processing Tools and is inspired by AMI-Net, an ML approach proposed for medical diagnosis with missing data in patient records.

The ALICE experiment was used as an R&D and testing environment; however, the proposed solution is general enough for other experiments with good PID capabilities (such as STAR at RHIC and others). Our approach improves the F1 score, a balanced measure of the PID purity and efficiency of the selected sample, for all investigated particle species (pions, kaons, protons).

[1] M. Kasak, K. Deja, M. Karwowska, M. Jakubowska, Ł. Graczykowski & M. Janik, "Machine-learning-based particle identification with missing data", Eur.Phys.J.C 84 (2024) 7, 691

[2] M. Karwowska, Ł. Graczykowski, K. Deja, M. Kasak, and M. Janik, "Particle identification with machine learning from incomplete data in the ALICE experiment", JINST 19 (2024) 07, C07013

References

[1] M. Kasak, K. Deja, M. Karwowska, M. Jakubowska, Ł. Graczykowski & M. Janik, "Machine-learning-based particle identification with missing data", Eur.Phys.J.C 84 (2024) 7, 691

[2] M. Karwowska, Ł. Graczykowski, K. Deja, M. Kasak, and M. Janik, "Particle identification with machine learning from incomplete data in the ALICE experiment", JINST 19 (2024) 07, C07013

Experiment context, if any

ALICE

Significance

At Warsaw University of Technology (WUT) physicists doing data analysis in ALICE have teamed up with researchers from other WUT faculties working in the domain of computer science and AI. As a leader of the WUT group ALICE, I have tried to find areas that are of interest to both fields, and PID is one of them. With the new PID approach, we aim to substitute the standard way of particle selection based on simple selections defined by users on the number of standard deviations for a given particle in a given detector. The presented work is an advanced method to approach the classification task in PID.

Author: Prof. GRACZYKOWSKI, Lukasz (Warsaw University of Technology (PL))

Co-authors: Dr DEJA, Kamil Rafal (Warsaw University of Technology (PL)); Ms KARWOWSKA, Maja (Warsaw University of Technology (PL)); Dr JANIK, Malgorzata Anna (Warsaw University of Technology (PL)); MYTKOWSKI, Marek Mateusz (Warsaw University of Technology (PL)); Mr KASAK, Milosz (Warsaw University of Technology (PL)); Dr JAKUBOWSKA, Monika Joanna (Warsaw University of Technology (PL))

Presenter: MYTKOWSKI, Marek Mateusz (Warsaw University of Technology (PL))

Session Classification: Poster session with coffee break

Track Classification: Track 2: Data Analysis - Algorithms and Tools