Contribution ID: **80** Type: **Oral**

Boosting statistical anomaly detection via multiple test with NPLM

Wednesday 13 March 2024 15:10 (20 minutes)

Statistical anomaly detection empowered by AI is a subject of growing interest at collider experiments, as it provides multidimensional and highly automatized solutions for signal-agnostic data quality monitoring, data validation and new physics searches.

AI-based anomaly detection techniques mainly rely on unsupervised or semi-supervised machine learning tasks. One of the most crucial and still unaddressed challenges of these applications is how to optimize the chances of detecting unexpected anomalies when prior knowledge about the nature of the latter is not available.

In this presentation we show how to exploit multiple tests to improve sensitivity to rare anomalies of different nature. We focus on a kernel methods based implementation of the NPLM algorithm, a signal-agnostic goodness of fit test based on a ML approximation of the likelihood ratio test [1, 2].

First, we show how performing multiple tests with different model configurations on the same data allows us to work around the problem of hyperparameters tuning, improving the algorithm's chance of discovery at the same time. Second, we show how multiple samples of streamed data can be optimally exploited to increase sensitivity to rare signals.

The presented findings offer the ability to perform fast, efficient, and sensitivity-enhanced applications of the NPLM algorithm to a larger and potentially more inclusive set of data, both offline and quasi-online.

With low-dimensional problems, we show this tool acts as a powerful diagnostic and compression algorithm. Furthermore, we find the agnostic nature of the strategy becomes especially relevant when the input data representation results from unsupervised ML algorithms, whose response to anomalies cannot be predicted.

Significance

The proposed strategies are new developments of the algorithm that have not been published yet. The tests carried out for this work show improved results over a set of benchmarks with respect to the previous implementation of the algorithm.

References

Previous work related to the topic:
https://link.springer.com/article/10.1140/epjc/s10052-022-10830-y
https://arxiv.org/abs/2305.14137
https://iopscience.iop.org/article/10.1088/2632-2153/acebb7

Experiment context, if any

CMS

Primary authors: Dr GROSSO, Gaia (IAIFI, MIT); Dr LETIZIA, Marco; HARRIS, Philip Coleman (Massachusetts Inst. of Technology (US))

Presenter: Dr GROSSO, Gaia (IAIFI, MIT)

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

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