

PHYSTAT-Anomalies

Report of Contributions

Contribution ID: 1

Type: **not specified**

Introduction

Tuesday, 24 May 2022 17:00 (10 minutes)

Presenters: LYONS, Louis (Imperial College (GB)); BEHNKE, Olaf (Deutsches Elektronen-Synchrotron (DE))

Session Classification: Session 1a. Chair: Mikael Kuusela

Contribution ID: 2

Type: **not specified**

Landscape of model independent searches

*Tuesday, 24 May 2022 17:10 (30 minutes)***ABSTRACT:**

This talk will briefly define and motivate anomaly detection at the Large Hadron Collider and will then give an overview of various method classifications based on the underlying physics goals and assumptions (and how these translate to statistical concepts). No one method will be able to cover all possibilities and it is essential to have a spectrum of techniques to achieve broad and deep model-independent sensitivity to physics beyond the Standard Model.

Presenter: NACHMAN, Ben (Lawrence Berkeley National Lab. (US))

Session Classification: Session 1a. Chair: Mikael Kuusela

Contribution ID: 3

Type: **not specified**

Learning New Physics from a machine

*Tuesday, 24 May 2022 17:55 (30 minutes)***ABSTRACT:**

Strategies to detect data departures from a given reference model, with no prior bias on the nature of the new physical model responsible for the discrepancy might play a vital role in experimental programs where, like at the LHC, increasingly rich experimental data are accompanied by an increasingly blurred theoretical guidance in their interpretation. I will describe one such strategy that employs neural networks, leveraging their virtues as flexible function approximants, but builds its foundations directly on the canonical likelihood-ratio approach to hypothesis testing. The algorithm compares observations with an auxiliary set of reference-distributed events, possibly obtained with a Monte Carlo event generator. It returns a p-value, which measures the compatibility of the reference model with the data. It also identifies the most discrepant phase-space region of the dataset, to be selected for further investigation. Imperfections due to mismodelling in the reference dataset can be taken into account straightforwardly as nuisance parameters. After illustrating the methodology, I will demonstrate its applicability to problems at a similar scale of complexity of realistic LHC analyses.

Presenter: WULZER, Andrea (Padova)**Session Classification:** Session 1a. Chair: Mikael Kuusela

Contribution ID: 4

Type: **not specified**

Relation of model independent searches with Goodness of Fit, 2-sample tests, etc.

*Tuesday, 24 May 2022 19:10 (30 minutes)***ABSTRACT:**

I will review goodness of fit testing and two sample testing in the context of trying to test for a new signal. My goal is to point to results in the statistics literature that might be unfamiliar in the physics community. Topics will include: optimal tests, classifier-based tests, reproducing kernel Hilbert space tests, level set tests, bump tests and robustness.

Presenters: WASSERMAN, Larry (Carnegie Mellon University); WASSERMAN, Larry (Carnegie Mellon University)

Session Classification: Session 1b. Chair Tom Junk

Contribution ID: 5

Type: **not specified**

Discussion

Tuesday, 24 May 2022 19:55 (30 minutes)

Presenters: COUSINS JR, Robert (University of California Los Angeles (US)); COUSINS JR, Robert (University of California Los Angeles (US))

Session Classification: Session 1b. Chair Tom Junk

Contribution ID: 6

Type: **not specified**

Challenges of anomaly detection with LHC data

Wednesday, 25 May 2022 17:00 (30 minutes)

ABSTRACT:

In recent years, there have been many proposed methodologies for machine learning anomaly detection at the LHC, such as those reported in the LHC Olympics and Dark Machines community reports. The first search using machine-learning anomaly detection was performed by ATLAS in the dijet final state, a fully data-driven analysis that uses the Classification Without Labels method and is complementary to the existing dedicated resonance searches.

In this talk, I will use the experience gained with the ATLAS weakly supervised dijet analysis to discuss the general

challenges of doing anomaly detection with LHC data and methodologies to address them.

Presenter: OCHOA, Ines (LIP Laboratorio de Instrumentacao e Fisica Experimental de Particulas (PT))

Session Classification: Session 2a: Chair: Ann Lee

Contribution ID: 7

Type: **not specified**

LHC Olympics

*Wednesday, 25 May 2022 17:45 (30 minutes)***ABSTRACT:**

We are at the beginning of a new era of data-driven, model-agnostic new physics searches at colliders that combine recent breakthroughs in anomaly detection and machine learning. This contribution will report on the LHC Olympics 2020, a community challenge accompanied by a set of simulated collider events. Participants in these Olympics have developed their methods using an R&D dataset and then tested them on black boxes: datasets with an unknown anomaly (or not). Methods made use of modern machine learning tools and were based on unsupervised learning (autoencoders, generative adversarial networks, normalizing flows), weakly supervised learning, and semi-supervised learning. We will review this challenge, including an overview of the competition, a description of methods deployed in the competition, lessons learned from the experience, and implications for data analyses with future datasets as well as future colliders.

Presenter: KASIECZKA, Gregor (Hamburg University)**Session Classification:** Session 2a: Chair: Ann Lee

Contribution ID: 8

Type: **not specified**

Supervised, unsupervised and data-derived signal regions

*Wednesday, 25 May 2022 19:00 (30 minutes)***ABSTRACT:**

Data-driven methods are becoming increasingly popular and could give us new insights when searching for signals from new physics. On the other hand, theoretical models and supervised learning approaches should not be neglected.

In this talk we present and compare different ways of defining “signal regions” at the Large Hadron Collider that are of interest for a “goodness-of-fit” test. We compare the performance of 3 of these approaches and discuss a new way to define “signal regions”.

Presenter: CARON, Sascha (Nikhef National institute for subatomic physics (NL))

Session Classification: Session 2b. Chair: David van Dyk

Contribution ID: 9

Type: **not specified**

Discussion

*Wednesday, 25 May 2022 19:45 (30 minutes)***Presenter:** LOCKHART, Richard (SFU)**Session Classification:** Session 2b. Chair: David van Dyk

Contribution ID: **10**Type: **not specified**

Closing remarks

*Wednesday, 25 May 2022 20:15 (5 minutes)***Presenter:** BEHNKE, Olaf (Deutsches Elektronen-Synchrotron (DE))**Session Classification:** Session 2b. Chair: David van Dyk

Contribution ID: **11**

Type: **not specified**

Discussion

Tuesday, 24 May 2022 18:40 (15 minutes)

Session Classification: Session 1a. Chair: Mikael Kuusela

Contribution ID: **12**Type: **not specified**

Discussion

*Wednesday, 25 May 2022 18:30 (15 minutes)***Session Classification:** Session 2a: Chair: Ann Lee