Anomaly detection in 5 mins or less

Snowmass 2022 CompF3 ML Session
ML for New Physics Searches

The vast majority of LHC searches for new physics are very model specific
ML for New Physics Searches

The vast majority of LHC searches for new physics are very model specific.

Why aren’t there more model-agnostic new physics searches?

Future Organization of Physics Analysis Groups at the LHC??

- B2G / HDBS
- SUSY
- Exotics/Exotica
- Search Groups
- Supporting organizations
- ML forum
- Statistics forum
- (Semi) Supervised
- Weakly Supervised
- Unsupervised
- Model Agnostic?
ML for Anomaly Detection

- How do we search for new physics in a model-agnostic way?
- Need two ingredients:
  1. Signal sensitivity (anomaly score)
  2. Background estimation

Both should be data-driven for a truly model-agnostic search
ML for Anomaly Detection

Types of Anomaly Scores

• Low p(x) — outliers

Autoencoders

Fully unsupervised

Farina, Nakai & DS 1808.08992
Heimel et al 1808.08979
and many more!!
ML for Anomaly Detection
Types of Anomaly Scores

- High $p_{data}(x)/p_{bg}(x)$ — overdensities

Enhanced bump hunts

Weakly supervised

CWoLa Hunting [Collins, Howe & Nachman 1805.02664, 1902.02634]
ANODE [Nachman & DS 2001.04990]
CATHODE [Hallin et al 2109.00546]
CURTAINS [Raine et al 2203.09470]
and more...
LHC Olympics 2020

The LHC Olympics 2020
A Community Challenge for Anomaly Detection in High Energy Physics


https://arxiv.org/abs/2101.08320
LHC Olympics 2020

3 “Black Box” datasets
LHC Olympics 2020

3 “Black Box” datasets

1. 1M QCD dijets + 834 $Z'\rightarrow XY$ signal (same topology as R&D, different masses)
3 “Black Box” datasets

1. 1M QCD dijets + 834 Z’->XY signal (same topology as R&D, different masses)
   - Several successful methods! (based on autoencoders, CWoLa, density estimation...)
LHC Olympics 2020

3 “Black Box” datasets

1. 1M QCD dijets + 834 Z’->XY signal (same topology as R&D, different masses)
   • Several successful methods! (based on autoencoders, CWoLa, density estimation…)

2. No signal
LHC Olympics 2020

3 “Black Box” datasets

1. 1M QCD dijets + 834 Z’->XY signal (same topology as R&D, different masses)
   • Several successful methods! (based on autoencoders, CWoLa, density estimation...)

2. No signal
   • Some approaches found false positives — importance and challenges of background estimation!
LHC Olympics 2020

3 “Black Box” datasets

1. 1M QCD dijets + 834 Z’->XY signal (same topology as R&D, different masses)
   • Several successful methods! (based on autoencoders, CWoLa, density estimation…)

2. No signal
   • Some approaches found false positives — importance and challenges of background estimation!

3. QCD dijets + 3,000 Z’ decaying to dijets or trijets
3 “Black Box” datasets

1. 1M QCD dijets + 834 Z’->XY signal (same topology as R&D, different masses)
   - Several successful methods! (based on autoencoders, CWoLa, density estimation…)

2. No signal
   - Some approaches found false positives — importance and challenges of background estimation!

3. QCD dijets + 3,000 Z’ decaying to dijets or trijets
   - No approaches discovered the signal in BB3
Outlook

• There is a lot of community interest in anomaly detection and model-agnostic NP searches!

• LHC Olympics 2020 was a very successful challenge, drawing nearly 50 participants from theory, experiment, and beyond (cosmology, computer science)

• Proofs-of-concept are beginning to be ported over to real data

  CWoLa Hunting
  ATLAS, PRL 125 131801 (2020)

  RNN VAE
  ATLAS-CONF-2022-045

• Many challenges for future R&D, including: feature selection, background estimation, multiple decay modes (BB3), non-resonant signals, …