

Tuesday, March 9, 2021 7:53 AM

2nd Terascale School: lecture on Anomaly Detection @ LHC "Model-Agnostic Searches for NP @ LHC"

Motivation & Context: We know there's NP out there beyond SM — e.g. dark matter!

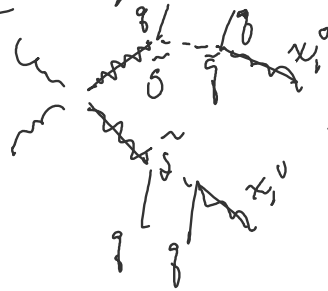
Yet still no sign of it @ LHC.

Maybe it's not there... however LHC has a huge dataset covering totally new phase space! enormous discovery potential!!

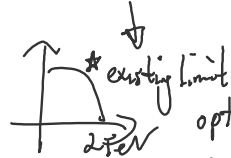
99.99% of searches

Example: SUSY search

gluon pair prod
of neutralino LSP



→ jets + MET



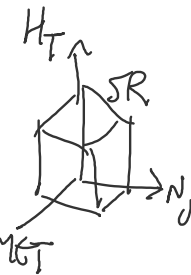
optimize cuts to
extend existing limit

defines SR

e.g. $N_j \geq 4$, $H_T \geq 4 \text{ TeV}$, $\text{MET} \geq \dots$ etc.

much remains to be explored

All but a few searches for NP @ LHC are highly model specific.



cont → N_{obs} vs. N_{bg}

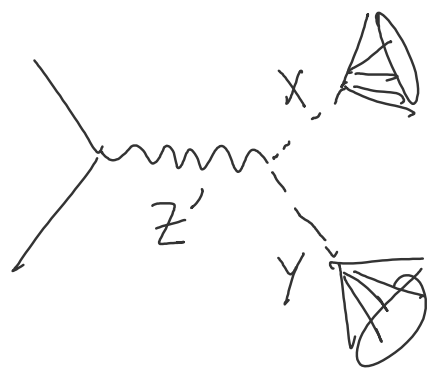
→ discovery or set limit

Tuesday, March 9, 2021 8:16 AM

These searches are worthwhile and should continue.

Yet it's ^{increasingly} important to consider more model-independent approaches to NP searches
b/c much phase space is not covered by model specific searches.

Example:



$X, Y \rightarrow \text{jets (e.g. } \bar{u}u \text{)}$

$m_X, m_Y \ll m_{Z'}$

$\Rightarrow X, Y$ are highly boosted.



⊙ R&D signal for LHC Olympics 2020

} No search @ LHC currently targeting it

Would be captured by inclusive dijet resonance search, but w/ greatly reduced sensitivity. (e.g. m_X, m_Y)

Tuesday, March 9, 2021 8:25 AM

There do exist a few model-agnostic searches @LHC:

1. $\begin{cases} \text{ATLAS General Search} \\ \text{CMS Multic} \end{cases} \xrightarrow{\text{IDEA}} \begin{array}{l} \text{Bin data into many categories (e.g. 2136 by MET)} \\ \text{Compare SM prediction derived from MC simulation} \end{array}$

Pros: completely signal model independent

Cons: huge look elsewhere effect (~ 1000 s or 1000 s of bins)

Heavily reliant on SM simulation!



"background model dependence"

SM simulations are far from perfect
can only discover NP if

$NP \Leftrightarrow \text{data} > \text{data} \Leftrightarrow \text{sim.}$

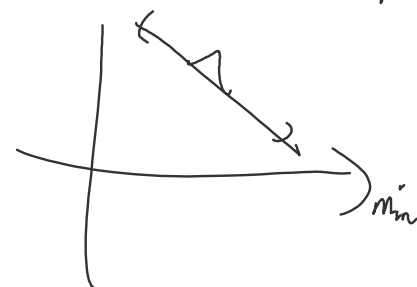
very challenging to completely characterize systematic uncertainties in this approach

1': Contour approach? looking for deviations b/w data & SM in measurement

Tuesday, March 9, 2021 8:33 AM

2. Bump hunts $\rightarrow$ assume signal is localized in some feature (usually m_{inv})
 not completely model independent $\rightarrow$ assume the b.g. is smooth in m_{inv} .
 but almost

Ex: dijet resonance search
 $d^2\text{photon}$ " "



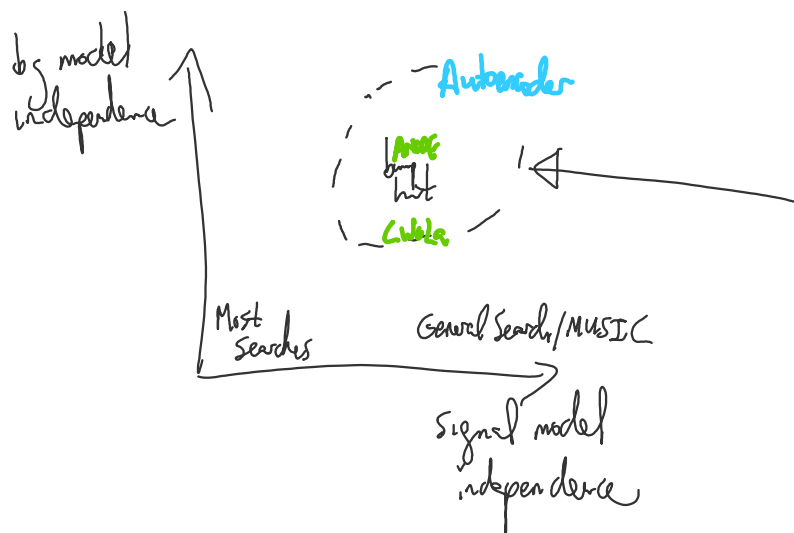
Pros: nearly completely signal & b.g. model independent
 enables b.g. estimation using sideband interpolation
 data-driven

Cons: inclusive versions might not be very sensitive if NP
 has additional features.
 or optimal

Tuesday, March 9, 2021 8:42 AM

Two general lessons so far:

1. Signal model independence & bg model indep are both important!
2. Two components to any anomaly search: discriminat (signal sensitivity)
bg prediction (background specificity)



In past few years, modern-ML-inspired methods have added a lot to this space!

Many new approaches were stimulated by LHC Olympics 2020
 ref. → Community paper just came out recently on arXiv.
 2101.08326
 many new approaches described there!

Tuesday, March 9, 2021 8:50 AM

A bit more on LHC0 2020

- Modeled after LHC olympics events ca. ~2008

↓
signal characterization

- No emphasis on anomaly detection

- 3 "black boxes" of 1M events (all hadronic) → asked people to say whether signal is in the BB's (p-value)

- R&D dataset of 1M QCD dijets

labeled

100k signal of type from before

$Z' \rightarrow XY, X, Y \rightarrow \bar{ij}$

in tutorial sess.
see also tutorial from
yesterday

- "Winter olympics" Jan 2020 @ ML4jets conference revealed BB1 — same topology as R&D but different masses

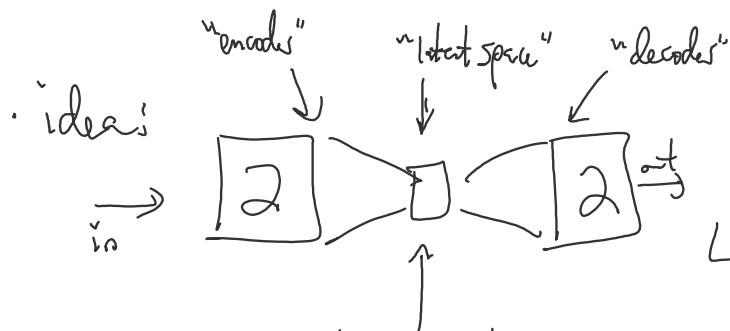
"Summer olympics" July 2020 revealed BB2 & BB3 → nobody got it
no signal $Z \rightarrow \bar{ij}$ or $\bar{ij}$

Tuesday, March 9, 2021 8:58 AM

3 New Approaches to Anomaly Detection @ UIC

1. Autoencoders Heibel, Kasiezko, Plehn, Thompson 1808.08979
 Farina, Nakai & DS 1808.08992

- widely used in "real world" anomaly detection (credit card fraud, self driving cars, terrorists, ...)
- dim'l reduction (data compression)
- generation



train ML algorithm to map input back to itself,
 $L = (\text{out} - \text{in})^2$ "MSE reconstruction error"

impose a bottleneck.
 lossy compression

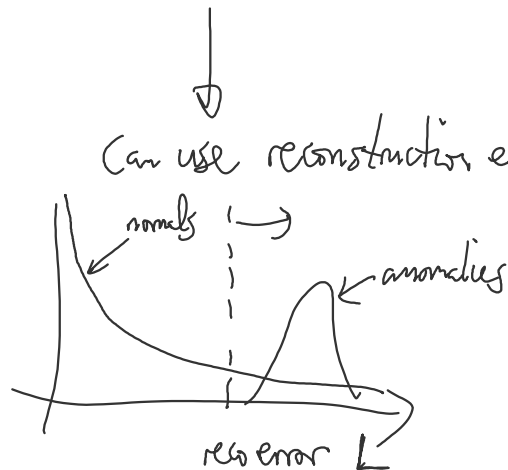
prevent AE from being identity map → forced to learn important underlying features of data

Tuesday, March 9, 2021 9:05 AM

For anomaly detection: train AE on "normal" events — learn to reconstruct them well
 learn a meaningful latent space

encounter "anomalous" events — AE doesn't know what to do
 reconstructs poorly.

can train
 directly on data
 if data is mostly
 "normal"



lots of details

- How to choose the cut?
- How to fix the AE architecture
- How to choose latent space size?

Pros: completely by & sig model independent

Cons: highly uncontrolled, may not catch every signal, likely to be very suboptimal

Tuesday, March 9, 2021 9:13 AM

2. & 3. : CWOLA Hunting & ANODE
 Collins, Hare & Nachman
 1805.02664
 Nachman & O'S
 2001.04990

Example:


 $m_i \quad m_{jj}$
 $\vec{x}: m_1, m_2, t_{21}^{(1)}, t_{21}^{(2)}$

- Diff. philosophy from AG — sacrifice a little model dependence for optimality
- Assume: signal is localized in some feature m while b_j is smooth $\rightarrow$ consider additional features $\vec{x}$ where signal could be localized.

- Optimal anomaly discriminant: Neyman Pearson Lemma $R(x) = \frac{p_{\text{data}}(x)}{p_{b_j}(x)}$

- CWOLA Hunting $\rightarrow$ use "likelihood ratio trick" — can likelihood ratio btw A & B $\leftrightarrow$ binary classifier on A vs. B.
- ANODE $\rightarrow$ use density estimator compute numerator & denominator separately.

Tuesday, March 9, 2021 9:20 AM

WOLA: optimal classifier on b_g & data $\rightarrow R(x) = \frac{P_{\text{data}}(x)}{P_{b_g}(x)}$

problem: we don't have labels in data for b_g !

~~$P_{b_g}(x)$~~

cf. general search
also uses $R(x)$

↓
limited version of LR
 $P_{b_g}(x)$ from simulation!

SB | SR | SB
n ↓

$$P(x | m \in \text{SB}) = P_{b_g}(x | m \in \text{SB}) = P_{b_g}(x | m \in \text{SR})$$

Assumption: x & m are uncorrelated
in b_g !!

Can train classifier on SR vs. SB!

$$R(x) = \frac{P(x | m \in \text{SR})}{P(x | m \in \text{SB})} = \frac{P_{\text{data}}^{\text{SR}}(x)}{P_{b_g}^{\text{SR}}(x)}$$

Tuesday, March 9, 2021 9:29 AM

ANODE: train conditional density estimator to learn $p(x|m \in SB)$, $p(x|m \in SR)$.



$$p_{\text{bg}}(x|m \in SB) \quad p_{\text{data}}(x|m \in SR)$$

integrate
automatically

$$p_{\text{bg}}(x|m \in SR)$$

construct explicitly $R(x)$.

- Pros: robust against x & m correlations
- Cons: classifier is easier to learn than density estimator
so if x & m uncorr., CWLb performance likely better.

