Novelty Detection in HEP Data Analysis

Ying-Ying Li

HKUST

April 16, 2019

Based on arXiv: 1807.10261
in collaboration with Jan Hajer, Tao Liu and He Wang

April 16, 2019

3rd IML Machine Learning Workshop
Novelty Detection in HEP Data Analysis

Supervised Learning

• High Energy Physics (HEP) is a big data science and has a long history of using supervised ML for data analysis: BDT, neural network, etc.
  a. neural network for top search @D0 (1990);
  b. MiniBooNe first used BDT and compared it with neural network (2004);
  c. BDT becomes more popular in HEP data analysis, e.g. in TOP2018, more than 50% of the results presented are based on BDT analysis.
• These supervised ML methods are highly efficient in analyzing signal events with complex topologies.
Novelty Detection in HEP Data Analysis

**Supervised Learning**

highly efficient in analysing signal events with complex topologies

**Q:**

- Given that new physics scenarios may share similar final states, can we search for them simultaneously and more efficiently

  Case I: di-top partner production vs Z’ production (decay to top pair)
  Case II: exotic Higgs decays (rich decay modes)

- Also, given the null results at LHC, new physics could be very unexpected.

- Supervised learning, being model-dependent, is incapable for these tasks.

**Novelty Detection**

Is “model”- independent, complementary to supervised learning, Allows us to detect new physics without a prior knowledge about it.
Novelty Detection in HEP Data Analysis

A review of novelty detection

Marco A.F. Pimentel *, David A. Clifton, Lei Clifton, Lionel Tarassenko

Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK

Abstract

Novelty detection is the task of classifying test data that differ in some respect from the data that are available during training. This may be seen as “one-class classification”, in which a model is constructed to describe “normal” training data. The novelty detection approach is typically used when the quantity of available “abnormal” data is insufficient to construct explicit models for non-normal classes. Application includes inference in datasets from critical systems, where the quantity of available normal data is very large, such that “normality” may be accurately modelled. In this review we aim to provide an updated and structured investigation of novelty detection research papers that have appeared in the machine learning literature during the last decade.
Novelty Detection in HEP Data Analysis

Workflow for Novelty Detection

- Step 1: (SM/background) feature learning
- Step 2: dimension reducing of feature space (auto-encoder)
- Step 3: novelty evaluating of testing data

Analyze detection sensitivity based on novelty response of testing data

Allow us to detect new physics model-independently

Novelty Detection in HEP Data Analysis

**Workflow for Novelty Detection**

1. **Step 1**: (SM/background) feature learning
2. **Step 2**: dimension reducing of feature space (auto-encoder)
3. **Step 3**: novelty evaluating of testing data

- Analyze detection sensitivity based on novelty response of testing data

---

Allow us to detect new physics model-independently

Novelty Evaluation

Applications:
- Electronic IT security
- Healthcare Informatics
- Medical Diagnostics and Monitoring
- Industrial monitoring & damage detection
- Text Mining
- Sensor Networks

Category:
- Probabilistic (2)
- Distance-based (3)
- Domain-based (4)
- Reconstruction-based (5)
- Information Theoretic (6)

Novelty Detection

nice review [M.A.F. Pimental et al (2014)]
Novelty Detection in HEP Data Analysis

Novelty Evaluation

Applications

- Electronic IT security
- Healthcare Informatics
- Medical Diagnostics and Monitoring
- Industrial monitoring & damage detection
- Text Mining
- Sensor Networks

Category

- Probabilistic (2)
- Distance-based (3)
- Domain-based (4)
- Reconstruction-based (5)
- Information Theoretic (6)

Novelty Detection

Novelty Detection in HEP Data Analysis

**Novelty Evaluators: traditional wisdom**

\[
\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d''_{\text{train}} \rangle^{1/2}}
\]

Novelty measure: range unnormalised

Novelty score: \(0 \leq \mathcal{O} \leq 1\)


[R.Socher, M.Ganjoo, C.D.Manning and A.Ng] (2013)
Novelty Evaluators: traditional wisdom

\[ \Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d''_{\text{train}} \rangle^{1/2}} \]

\[ \mathcal{O} = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{c\Delta}{\sqrt{2}} \right) \right) \]

- However, this design is insensitive to the clustering of the testing data with unknown pattern
- Recall: the clustering features such as resonances, shape, etc., could be important for BSM physics detection
- The testing data of unknown pattern with such features are scored low, unless they are away from the training data distribution!
Novelty Detection in HEP Data Analysis

New Novelty Measure

\[
\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}}
\]

\[
\Delta_{\text{new}} = \frac{d^{-m}_{\text{test}} - d^{-m}_{\text{train}}}{d^{-m/2}_{\text{train}}}
\]

- \(d_{\text{train}}\) mean distance of a testing data point to its k nearest neighbors in the training dataset
- \(d_{\text{test}}\) mean distance of a testing data point to its k nearest neighbors in the testing dataset
- \(m\) dimension of the feature space

Novelty response is evaluated by comparing local densities in the training and testing datasets.

Approximately statistical interpretation: \(\Delta_{\text{new}} \propto \frac{S}{\sqrt{B}}\) local bin

Novelty Detection in HEP Data Analysis

New Novelty Measure

\[ \Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d''_{\text{train}} \rangle^{1/2}} \]

\[ \Delta_{\text{new}} = \frac{d_{\text{test}} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}} \]

- Big density difference => high score
- Small density difference => low score
- => a measure of clustering
Novelty Detection in HEP Data Analysis

New Novelty Measure

\[
\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d''_{\text{train}} \rangle^{1/2}}
\]

\[
\Delta_{\text{new}} = \frac{d^{-m}_{\text{test}} - d^{-m}_{\text{train}}}{d^{-m/2}_{\text{train}}}
\]

- Consider 2D Gaussian samples
- Training dataset: known pattern only
- Testing dataset: known + unknown patterns
- Compared to O_trad, the novelty response of unknown-pattern data is much stronger for O_new
- \( \Rightarrow \) A well-separation between the known- and unknown-pattern data distributions
Novelty Detection in HEP Data Analysis

Addressing Look Elsewhere Effect

Without a priori knowledge on the BSM physics, novelty detection might suffer from a large "Look Elsewhere Effect (LEE)'', given the feature space to probe!

\[ \Delta_{\text{new}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}} \]

Novelty Detection in HEP Data Analysis

Addressing Look Elsewhere Effect

\[ \Delta_{\text{new}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}} \]


The influence of fluctuations for detection sensitivity can be compensated for as the luminosity \( L \) increases, if \( k \) scales with \( L \).

Central Limit Theorem: the standard deviation of Delta_new scales with 1/sqrt{\( k \)} or 1/sqrt{\( L \)}, for the testing data with known patterns only.
Novelty Detection in HEP Data Analysis

Addressing Look Elsewhere Effect

\[ \Delta_{\text{new}} = \frac{d_{\text{test}} - d_{\text{train}}}{d_{\text{train}}^{m/2}} \]

The influence of fluctuations for detection sensitivity can be compensated for as the luminosity \( L \) increases, if \( k \) scales with \( L \).

Central Limit Theorem: the standard deviation of \( \Delta_{\text{new}} \) scales with 1/\( \sqrt{k} \) or 1/\( \sqrt{L} \), for the testing data with known patterns only.

Suppression by luminosity might not be sufficient
Addressing Look Elsewhere Effect

Addressing Look Elsewhere Effect

B and C: more data in the non-signal region.

Fluctuations

Novelty Detection in HEP Data Analysis

Addressing Look Elsewhere Effect

To compensate for high-scoring (by $O_{\text{new}}$) of known-pattern data from high-density region

$O_{\text{comb}} = \sqrt{O_{\text{trad}} O_{\text{new}}}$

B and C: more data in the non-signal region.
Addressing Look Elsewhere Effect


Center slightly shifted, with S/B=1/20

O_comb based analysis yields more than 50% improvement in detection sensitivity!
Benchmark Analysis

Analysis one: di-top(leptonic) production at LHC

- \( pp \to \bar{t}_l t_l \) , \( \sigma = 11.5 \text{ fb} \), \( X_1: pp \to \bar{T}T \to W_l^+ W_l^- b \bar{b} \)
- \( pp \to t_l \bar{b} W_l^{\pm} \) , \( \sigma = 0.365 \text{ fb} \), \( X_2: pp \to Z' \to \bar{t}t \)
- \( pp \to Z_b Z_l \) , \( \sigma = 0.0765 \text{ fb} \).

Analysis two: exotic Higgs decays at e+e- collider

- \( e^+ e^- \to hZ \to Z_{inv}^* Z_{bb} l^+ l^- \) \( \sigma = 0.00686 \text{ fb} \), \( Y_1: h \to \tilde{\chi}_1 \tilde{\chi}_2 \to \tilde{\chi}_1 \tilde{\chi}_1 a. \)
- \( e^+ e^- \to hZ \to Z_{bb}^* Z_{inv} l^+ l^- \) \( \sigma = 0.00259 \text{ fb} \). \( Y_2: h \to Z a \)

<table>
<thead>
<tr>
<th>Parameter values</th>
<th>( \sigma (\text{fb}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( X_1 ) ( m_T = m_{\bar{T}} ) 1.2 TeV, ( \text{BR}(T \to W_l^+ b) = 50% )</td>
<td>0.152</td>
</tr>
<tr>
<td>( X_2 ) ( m_{Z'} = 3 \text{ TeV}, g_{Z'} = g_Z, \text{BR}(Z' \to tt) = 16.7% )</td>
<td>1.55</td>
</tr>
<tr>
<td>( Y_1 ) ( m_{N_1} = \frac{m_{N_2}}{g} = \frac{m_{\alpha}}{4} = 10 \text{ GeV}, \text{BR}(h \to bbE_T^{miss}) = 1% )</td>
<td>0.108</td>
</tr>
<tr>
<td>( Y_2 ) ( m_{\alpha} = 25 \text{ GeV}, \text{BR}(h \to bbE_T^{miss}) = 1% )</td>
<td>0.053</td>
</tr>
</tbody>
</table>
Novelty Detection in HEP Data Analysis

Benchmark Analysis


- X1: well-modelled by the Gaussian sample!
- X2: O_comb less efficient due to one-order larger S/B
- Y1 and Y2: O_new performs universally better than the others, due to large S/B
- The sensitivities based on the algorithm designed are not far below the ones set by supervised learning
Novelty Detection in HEP Data Analysis

Conclusion and Outlook

- Proposed workflow for novelty detection
- New novelty evaluators are proposed to address the clustering and LEE effect;
  
  Follow-up study to parton showers and full detector simulations.

Interesting questions to study:

- Understand which features are captured when auto-encoder doing dimension reduction;
- Invent a novelty evaluator to exploit multiple measures at once

... in collaboration with Jiang, Juste Rozas, Liu and Tu
We are heading forward to the truth of our nature
gradually and passionately

Thank YOU!