Masters Presentation

Latent Variable Machine Learning Algorithms: Applications in a Nuclear Physics Experiment

Robert Solli

University of Oslo, Expert Analytics AS

November 12, 2019

Outline

1. Introducing the Active Target Time Projection Chamber (AT-TPC)
2. Challenges with traditional analysis of AT-TPC data - and the evangilization of Machine Learning
   (i) Recap of central literature
   (ii) Introducing thesis problem statements
3. An introduction to central Machine Learning concepts
4. The Auto-Encoder neural network
5. Results
6. Discussion, Summary, Conclusion and Outlook

AT-TPC

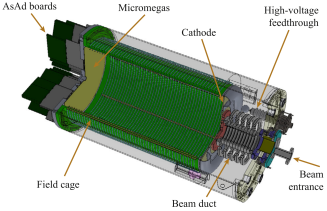


Figure: Diagram of the AT-TPC[1]

The AT-TPC is an experiment set up at the rare isotopes facility on the Michigan State University campus. The AT-TPC is commissioned to capture reactions with exotic nuclei.

[1] J. Bradt et al. "Commissioning of the Active-Target Time Projection Chamber". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 875 (2017), p. 65. DOI: 10.1016/J.NIMA.2017.09.013. URL: https://www.sciencedirect.com/science/article/pii/S0168900217309683.

AT-TPC

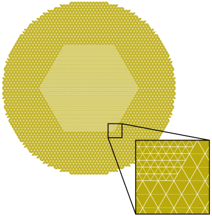


Figure: Detector pad plane of the AT-TPC[2]

Each triangle represents spatial discrete regions of the detector surface. The pad-plane consists of some 10^4 sensor pads on a circle with $r = 29cm$

[2] Bradt et al., "Commissioning of the Active-Target Time Projection Chamber".

AT-TPC Data

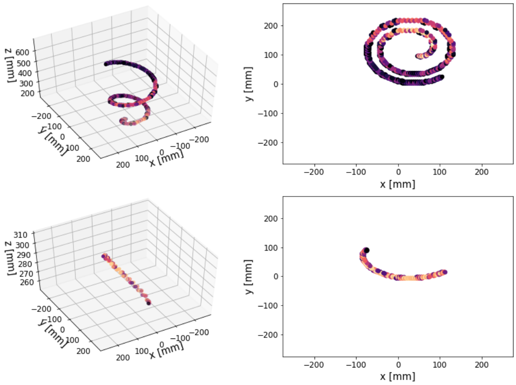
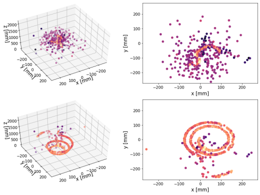
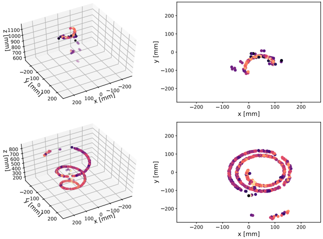


Figure: Simulated AT-TPC data in an experiment with ^{46}Ar

AT-TPC Data



(a) Unfiltered events from the ^{46}Ar experiment



(b) Noise-filtered events from the ^{46}Ar experiment

Significant noise levels in the experiment, from unknown sources.
Some 60% of the recorded events are from unidentified reactions.

Challenges with the AT-TPC

I Expensive integration - for each event a fit is computed

II Assumptions of the integration technique:

  (i) Each event is fit against parameters of the event of interest

  (ii) The integration is sensitive to Noise and Breaks in the tracks

The amount of data is also significant: the experiment generates on the order of 10^5 events per hour running.

Idea

Solve the problem by training deep neural networks, a very flexible algorithm from the Machine Learning community.

Previous Work

- Work on applying ML to this data started with a supervised learning project by Kuchera et al.[3].
- The authors explored a *supervised classification* problem of identifying reactions when ground-truth labels available.
- By fine tuning pre-trained networks the authors achieve very impressive performance.

One of the open questions is then, can we segment the events based on reactions without the ground truth labels?

[3] Michelle. P. Kuchera et al. "Machine learning methods for track classification in the AT-TPC". In: *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 940 (2019), p. 156. DOI: 10.1016/j.nima.2019.05.097. URL: https://linkinghub.elsevier.com/retrieve/pii/S0168900219308046.

ML background

Machine Learning (ML) is an amorphous set of algorithms for pattern recognition and function approximation. Bred as a mix between computer science and statistical learning theory, it includes algorithms like:

I. Linear Regression
II. Logistic Regression
III. Random Forest Classifiers
IV. Genetic Algorithms

And many others...

Deep Learning

A special sub-branch of Machine Learning is the field of Deep Learning, or more broadly: Differentiable Programming

The premise of Deep Learning is to formulate an approximation to some unknown function $f(\mathbf{x})$ with a model, $\hat{f}$, that maintains some set of parameters $\{\theta\}$.

Some examples of the unknown function, $f(\mathbf{x})$, we would want to approximate are:

(a) a Hamiltonian of a system

(b) a function which determines the thermodynamic phase of a system

(c) a function which itentifies dog-species from a picture

Deep Learning

Then then model can be tuned with gradient methods, the simplest of which is a steepest descent update:

$$\theta_i \leftarrow \theta_i - \eta \frac{\partial \mathcal{C}(\mathbf{x}, f, \hat{f})}{\partial \theta_i}, \tag{1}$$

moderated by a *learning rate* η to ensure that the steps are small enough.
The functional $\mathcal{C}$ is the *cost* function for the problem and is commonly a variation of either, the Mean Squared Error:

$$\mathcal{C}(\mathbf{x}, f, \hat{f}) = \sum (f(\mathbf{x})_i - \hat{f}(\mathbf{x})_i)^2, \tag{2}$$

or the Cross Entropy

$$\mathcal{C}(\mathbf{x}, f, \hat{f}) = -\sum f(\mathbf{x})_i \log \hat{f}(\mathbf{x})_i. \tag{3}$$

Deep Learning: Neural Networks

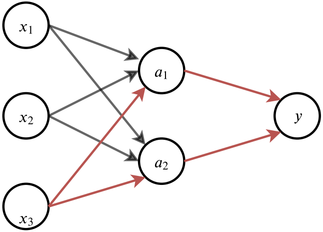


Figure: A neural network with three input nodes, two hidden nodes, and one output node

Each of the *activations* **a** are computed as a matrix product fed through a nonlinear activation *activation function* g

$$\boldsymbol{a}^{[1]} = g(\boldsymbol{x}\boldsymbol{\theta}^{[1]})_D \tag{4}$$

Autoencoders

- Recall that we want to separate classes of reaction products
- Additionally - we assume that we have access to very little or no ground truth labelled data

Idea

Learn the distribution over the events through two nonlinear maps which compress and inflate a representation of the events.

Central Hypothesis

If we can compress a representation of an event reconstruct it - the compression must be informative of the type of event.

Autoencoders

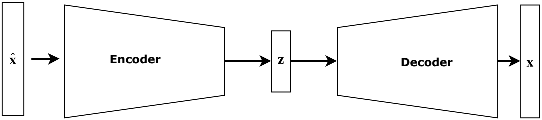


Figure: Autoencoder neural network schematic

An autoencoder is defined by an encoder/decoder pair of neural networks. We construct these such that

$$\dim(\hat{\mathbf{x}}) \gg \dim(\mathbf{z}), \tag{5}$$

and with the optimization objective

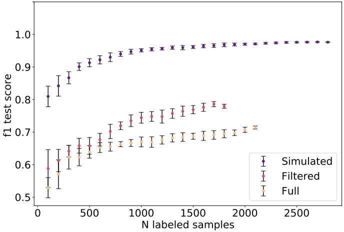
$$\mathcal{O} = \arg\min ||\hat{\mathbf{x}} - \mathrm{Decoder}(\mathrm{Encoder}(\hat{\mathbf{x}}))||_2^2. \tag{6}$$

Experiment

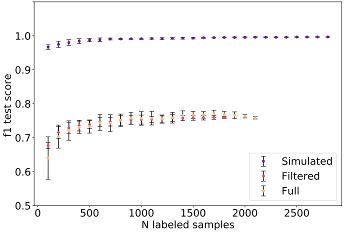
- We chose to represent the data as 2D projections, neglecting the z-axis.
- An autoencoder network is then fit to the events in an end to end manner.
- After traing, we extract the compressed (or *latent*) expressions of the data.
- With varying amounts of labelled ground-truth data we fit a very simple classifier to the latent expressions

We wish to construct latent spaces that are as close to trivially separable, and so use a logistic regression classifier.

Results



(a) Performance for an autoencoder with a naive architecture



(b) Performance using a representation of the events as seen by a VGG16 network

$$\text{recall} = \frac{TP}{TP + FP}, \quad \text{precision} = \frac{TP}{TP + FN}. \tag{7}$$

The $f1$ score is defined as the harmonic mean of precision and recall for each class. Formally we define it as

$$f1 = 2\frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}. \tag{8}$$