Machine Learning Overview

Tilman Plehn

Universität Heidelberg

Wits/Liverpool/Zoom 3/2021

Nothing is ever new

LHC visionaries

– 1991: NN-based quark-gluon tagger [visionary: Lönnblad, Peterson, Rögnvaldsson]

USING NEURAL NETWORKS TO IDENTIFY JETS

Leif LÖNNBLAD*, Carsten PETERSON** and Thorsteinn RÖGNVALDSSON***

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 29 June 1990

A neural network method for identifying the ancestor of a hadron jet is presented. The idea is to find an efficient mapping between certain observed hadronic kinematical variables and the quark-gluon identity. This is done with a neuronic expansion in terms of a network of sigmoidal functions using a gradient descent procedure, where the errors are back-propagated through the network. With this method we are able to separate gluon from quark jets originating from Monte Carlo generated e^+e^- events with $\sim 85\%$ approach. The result is independent of the MC model used. This approach for isolating the gluon jet is then used to study the so-called string effect.

In addition, heavy quarks (b and c) in e^+e^- reactions can be identified on the 50% level by just observing the hadrons. In particular we are able to separate b-quarks with an efficiency and purity, which is comparable with what is expected from vertex detectors. We also speculate on how the neural network method can be used to disentangle different hadronization schemes by compressing the dimensionality of the state space of hadrons.

Nothing is ever new

LHC visionaries

– 1991: NN-based quark-gluon tagger [visionary: Lönnblad, Peterson, Rögnvaldsson]

– 1994: jet-algorithm W/top-tagger [Seymour]

USING NEURAL NETWORKS TO IDENTIFY JETS

Leif LÖNNBLAD*, Carsten PETERSON** and Thorsteinn RÖGNVALDSSON***

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 29 June 1990

A neural network method for identifying the ancestor of a hadron jet is presented. The idea is to find an efficient mapping between certain observed hadronic kinematical variables and the quark-gluon identity. This is done with a neuronic expansion in terms of a network of sigmoidal functions using a gradient descen network. With this method we ar Carlo generated e^+e^- events v model used. This approach for i effect.

In addition, heavy quarks (b a just observing the hadrons. In pa purity, which is comparable with how the neural network method compressing the dimensionality c

Searches for new particles using cone and cluster jet algorithms: a comparative study

Michael H. Seymour

Department of Theoretical Physics, University of Lund, Sölvegatan 14A, S-22362 Lund, Sweden

Received 18 June 1993; in revised form 16 September 1993

Abstract. We discuss the reconstruction of the hadronic decays of heavy particles using jet algorithms. The ability to reconstruct the mass of the decaying particle is compared between a traditional cone-type algorithm and a recently proposed cluster-type algorithm. The specific examples considered are the semileptonic decays of a heavy Higgs boson at $\sqrt{s}=16$ TeV, and of top quark–antiquark pairs at $\sqrt{s}=1.8$ TeV. We find that the cluster algorithm offers considerable advantages in the former case, and a slight advantage in the latter. We briefly discuss the effects of calorimeter energy resolution, and show that a typical resolution dilutes these advantages, but does not remove them entirely.

except that the invariant mass of a pair is replaced by the transverse momentum of the softer particle relative to the other.

More recently, this algorithm was extended to collisions with incoming hadrons [5], and a longitudinally-invariant $k_\perp$-clustering algorithm for hadron-hadron collisions was proposed [6]. This algorithm has been compared with the more commonly used cone algorithm from the viewpoints of a parton-shower Monte Carlo program [6, 7], and a fixed-order matrix-element calculation [8], and advantages of the cluster algorithm were reported in both cases. This paper is concerned with a comparison between the algorithms for the task of reconstructing the hadronic decays of heavy particles, which was also studied in a preliminary way in [9].

The only as-yet unobserved particles of the minimal Standard Model are the top quark and Higgs boson. The search for, and study of, these particles are among the most important goals of current and planned hadron-

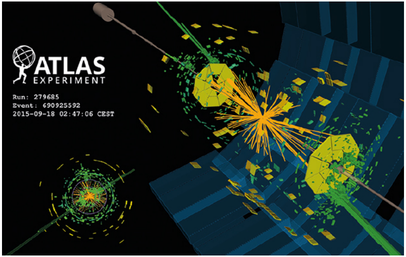
Why LHC jets?

Data from ATLAS & CMS

– most LHC interactions $q\bar{q}, gg \to q\bar{q}, gg$

– quarks/gluon visible as jets $\quad \sigma_{pp\to jj} \times \mathcal{L} \approx 10^8\text{fb} \times 80/\text{fb} \approx 10^{10}$ events

$\Rightarrow$ Tons of data



Why LHC jets?

Data from ATLAS & CMS

– most LHC interactions $q\bar{q}, gg \to q\bar{q}, gg$

– quarks/gluon visible as jets $\quad \sigma_{pp\to jj} \times \mathcal{L} \approx 10^8\text{fb} \times 80/\text{fb} \approx 10^{10}$ events

$\Rightarrow$ Tons of data

Physics in jets

– re-summed perturbative QFT prediction from QCD

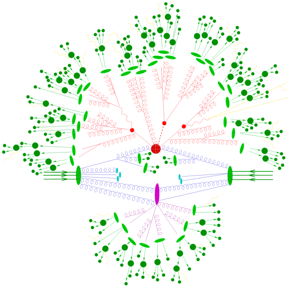
– jets as decay products

67% $W \to jj$ $\quad$ 70% $Z \to jj$ $\quad$ 60% $H \to jj$ $\quad$ 67% $t \to jjj$ $\quad$ 60% $\tau \to j$ ...

– flavor tagging classic multivariate

– new physics in ‘dark showers’

$\Rightarrow$ Fundamentally interesting



Why LHC jets?

Data from ATLAS & CMS

– most LHC interactions $q\bar{q}, gg \to q\bar{q}, gg$

– quarks/gluon visible as jets $\sigma_{pp\to jj} \times \mathcal{L} \approx 10^8\text{fb} \times 80/\text{fb} \approx 10^{10}$ events

$\Rightarrow$ Tons of data

Physics in jets

– re-summed perturbative QFT prediction from QCD

– jets as decay products

67% $W \to jj$ 70% $Z \to jj$ 60% $H \to jj$ 67% $t \to jjj$ 60% $\tau \to j$ ...

– flavor tagging classic multivariate

– new physics in ‘dark showers’

$\Rightarrow$ Fundamentally interesting

Subjets for the cool stuff

– resonance searches in VV, VH, $t\bar{t}$

– target masses high
target EFT kinematic same

$\Rightarrow$ Why invent high-level observables?

Early ML-years

A brief history of ML-tagging

– 2014/15: first jet image papers [Cogan, Kagan, Strauss, Schwartzman, de Oliveira, Mackey, Nachman]

– 2017: first (working) ML top tagger [Kasieczka, TP, Russell, Schell]

– ML4Jets 2017: what architecture works best?

To see how our DeepTopLoLa tagger deals with this problem and to test what kind of structures drive the network output, we turn the problem around and ask the question if the Minkowski metric is really the feature distinguishing top decays and QCD jets. To this end, we define the invariant mass $m(\tilde{k}_j)$ and the distance d^2_{jm} in Eq.(6) with a trainable diagonal metric. After applying a global normalization we find

$$g = \text{diag}(\quad 0.99 \pm 0.02, \\ -1.01 \pm 0.01, -1.01 \pm 0.02, -0.99 \pm 0.02) \, , \tag{9}$$

where the errors are given by five independently trained copies. It is crucial for our physics understanding [37] that the distinguishing power of the DeepTopLoLa tagger is indeed the same mass drop [1] that drives many QCD-based top taggers [6,7] and the image-based top tagger, as shown in detail in Ref. [20].

Machine Learning

Tilman Plehn

Jet tagging

Anomalies

Simulaton

Inversion

Inference

Early ML-years

A brief history of ML-tagging

- 2014/15: first jet image papers [Cogan, Kagan, Strauss, Schwartzman, de Oliveira, Mackey, Nachman]
- 2017: first (working) ML top tagger [Kasieczka, TP, Russell, Schell]
- ML4Jets 2017: what architecture works best?
- ML4Jets 2018: lots of architectures work [1902.09914]

SciPost Physics — Submission

The Machine Learning Landscape of Top Taggers

G. Kasieczka (ed)[1], T. Plehn (ed)[2], A. Butter[2], K. Cranmer[3], D. Debnath[4], B. M. Dillon[5], M. Fairbairn[6], D. A. Faroughy[5], W. Fedorko[7], C. Gay[7], L. Gouskos[8], J. F. Kamenik[5,9], P. T. Komiske[10], S. Leiss[1], A. Lister[7], S. Macaluso[3,4], E. M. Metodiev[10], L. Moore[11], B. Nachman,[12,13], K. Nordström[14,15], J. Pearkes[7], H. Qu[8], Y. Rath[16], M. Rieger[16], D. Shih[4], J. M. Thompson[2], and S. Varma[6]

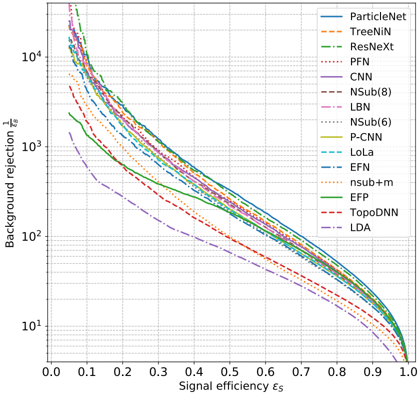
1 Institut für Experimentalphysik, Universität Hamburg, Germany
2 Institut für Theoretische Physik, Universität Heidelberg, Germany
3 Center for Cosmology and Particle Physics and Center for Data Science, NYU, USA
4 NHECT, Dept. of Physics and Astronomy, Rutgers, The State University of NJ, USA
5 Jozef Stefan Institute, Ljubljana, Slovenia
6 Theoretical Particle Physics and Cosmology, King's College London, United Kingdom
7 Department of Physics and Astronomy, The University of British Columbia, Canada
8 Department of Physics, University of California, Santa Barbara, USA
9 Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia
10 Center for Theoretical Physics, MIT, Cambridge, USA
11 CP3, Universitéxx Catholique de Louvain, Louvain-la-Neuve, Belgium
12 Physics Division, Lawrence Berkeley National Laboratory, Berkeley, USA
13 Simons Inst. for the Theory of Computing, University of California, Berkeley, USA
14 National Institute for Subatomic Physics (NIKHEF), Amsterdam, Netherlands
15 LPTHE, CNRS & Sorbonne Université, Paris, France
16 III. Physics Institute A, RWTH Aachen University, Germany

gregor.kasieczka@uni-hamburg.de
plehn@uni-heidelberg.de

July 24, 2019

Abstract

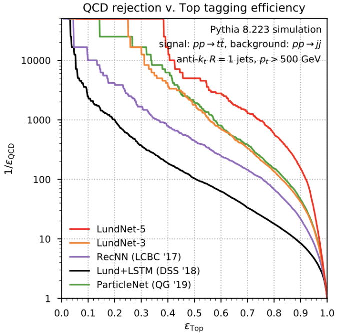
Based on the established task of identifying boosted, hadronically decaying top quarks, we compare a wide range of modern machine learning approaches. Unlike most established methods they rely on low-level input, for instance calorimeter output. While their network architectures are vastly different, their performance is comparatively similar. In general, we find that these new approaches are extremely powerful and great fun.



Early ML-years

A brief history of ML-tagging

– 2014/15: first jet image papers [Cogan, Kagan, Strauss, Schwartzman, de Oliveira, Mackey, Nachman]
– 2017: first (working) ML top tagger [Kasieczka, TP, Russell, Schell]
– ML4Jets 2017: what architecture works best?
– ML4Jets 2018: lots of architectures work [1902.09914]
– ML4Jets 2020: nothing is ever over [Dreyer, Carrazza, Qu]



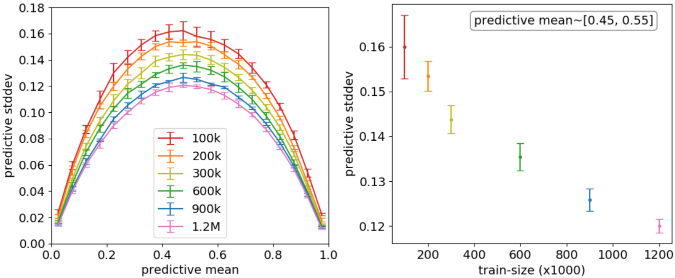
Early ML-years

A brief history of ML-tagging

- 2014/15: first jet image papers [Cogan, Kagan, Strauss, Schwartzman, de Oliveira, Mackey, Nachman]
- 2017: first (working) ML top tagger [Kasieczka, TP, Russell, Schell]
- ML4Jets 2017: what architecture works best?
- ML4Jets 2018: lots of architectures work [1902.09914]
- ML4Jets 2020: nothing is ever over [Dreyer, Carrazza, Qu]

Tagging with per-jet errors [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson]

- Bayesian tagging network
- similar performance as deterministic network
- Per-jet error: training statistics



Anomaly searches

From supervised to unsupervised learning

- what ML people consider cool
- knowledge just an unwanted bias
  elevate ignorance to structural requirement
- unsupervised learning/anomaly searches [remember Bruce Knutsen?]
- fun LHC applications: Tao Liu's talk

Novelty Detection Meets Collider Physics

Jan Hajer,[1,2] Ying-Ying Li,[3,4] Tao Liu,[3] and He Wang[3]

[1] *Institute for Advanced Studies, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong S.A.R, P.R.China*
[2] *Centre for Cosmology, Particle Physics and Phenomenology, Université catholique de Louvain, Louvain-la-Neuve B-1348, Belgium*
[3] *Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong S.A.R., P.R.China*
[4] *Kavli Institute for Theoretical Physics, University of California Santa Barbara, CA 93106–4030, USA*

Novelty detection is the machine learning task to recognize data, which belong to an unknown pattern. Complementary to supervised learning, it allows to analyze data model-independently. We demonstrate the potential role of novelty detection in collider physics, using autoencoder-based deep neural network. Explicitly, we develop a set of density-based novelty evaluators, which are sensitive to the clustering of unknown-pattern testing data or new-physics signal events, for the design of detection algorithms. We also explore the influence of the known-pattern data fluctuations, arising from non-signal regions, on detection sensitivity. Strategies to address it are proposed. The algorithms are applied to detecting fermionic di-top partner and resonant di-top productions at LHC, and exotic Higgs decays of two specific modes at a future e^+e^- collider. With parton-level analysis, we conclude that potentially the new-physics benchmarks can be recognized with high efficiency.

ML in simulation

Fundamental understanding a unique LHC feature

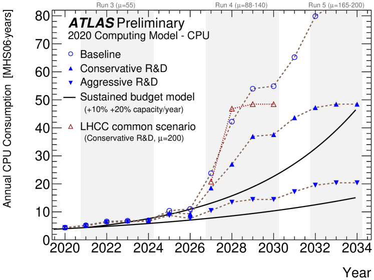
- precision theory
- precision simulations
- precision measurements

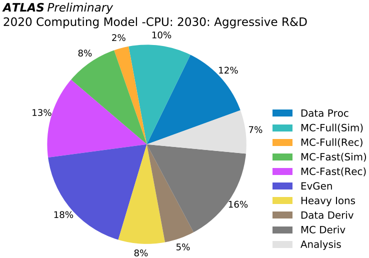
ML in simulation

Fundamental understanding a unique LHC feature

– precision theory
– precision simulations
– precision measurements

⇒ What's needed to keep the edge?





ML in simulation

Fundamental understanding a unique LHC feature

- precision theory
- precision simulations
- precision measurements

⇒ What's needed to keep the edge?

Precision event generation

- simulated event numbers ∼ expected events [factor 25 for HL-LHC]
- general move to NLO/NNLO [1%-2% error]
- higher relevant multiplicities [jet recoil, extra jets, WBF, etc.]
- new low-rate high-multiplicity backgrounds
- cutting-edge predictions not through generators [N^3LO in Pythia?]
- interpretation beyond specific models [jets+MET]

ML in simulation

Fundamental understanding a unique LHC feature

– precision theory
– precision simulations
– precision measurements
⇒ What's needed to keep the edge?

Precision event generation

– simulated event numbers ∼ expected events [factor 25 for HL-LHC]
– general move to NLO/NNLO [1%-2% error]
– higher relevant multiplicities [jet recoil, extra jets, WBF, etc.]
– new low-rate high-multiplicity backgrounds
– cutting-edge predictions not through generators [N^3LO in Pythia?]
– interpretation beyond specific models [jets+MET]

Three ways to use ML

– improve current tools: iSherpa, ML-MadGraph, etc
– new ideas, like fast ML-generator-networks
– conceptual ideas in theory simulations and analyses

Coolest ML-algorithm

Generative adversarial network

- training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
- discriminator constructing $D(x)$ by minimizing [classifier $D(x) = 1, 0$ true/generator]

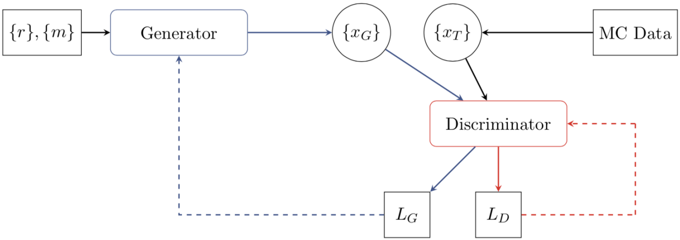
$$L_D = \big\langle -\log D(x) \big\rangle_{x_T} + \big\langle -\log(1 - D(x)) \big\rangle_{x_G}$$

- generator constructing $r \to x_G$ by minimizing [D needed]

$$L_G = \big\langle -\log D(x) \big\rangle_{x_G}$$

- equilibrium $D = 0.5 \;\Rightarrow\; L_D = L_G = -\log 0.5$

$\Rightarrow$ statistically independent copy of training events



Coolest ML-algorithm

Generative adversarial network

– training: true events $\{x_T\}$
  output: generated events $\{r\} \to \{x_G\}$
– discriminator constructing $D(x)$ by minimizing [classifier $D(x) = 1, 0$ true/generator]
– generator constructing $r \to x_G$ by minimizing [D needed]

$\Rightarrow$ statistically independent copy of training events

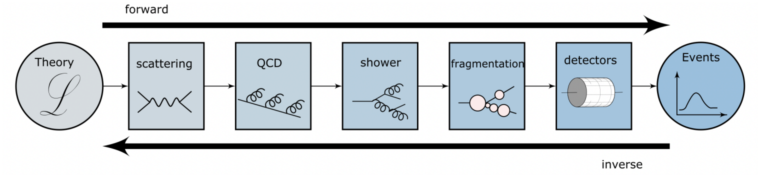
Vast number of studies

– Jets [de Oliveira (2017), Carrazza-Dreyer (2019)]
– Detector simulations [Paganini (2017), Musella (2018), Erdmann (2018), Ghosh (2018), Buhmann (2020)]
– Events [Otten (2019), Hashemi, DiSipio, Butter (2019), Martinez (2019), Alanazi (2020), Chen (2020), Kansal (2020)]
– Unfolding [Datta (2018), Omnifold (2019), Bellagente (2019), Bellagente (2020), Howard (2020)]
– Templates for QCD factorization [Lin (2019)]
– EFT models [Erbin (2018)]
– Event subtraction [Butter (2019)]
– Sherpa [Bothmann (2020), Gao (2020)]
– Basics [GANplification (2020), DCTR (2020)]
– Unweighting [Verheyen (2020), Backes (2020)]
– Superresolution [DiBello (2020), Baldi (2020)]

Inversion

Beyond forward simulation [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

– bijective transformation — physics mapped on latent space
  Jacobian tractable — normalizing flow [specifically: coupling layer]
  evaluation in both directions — INN [Ardizzone, Rother, Köthe]
– conditional GAN/INN: inverted events generated



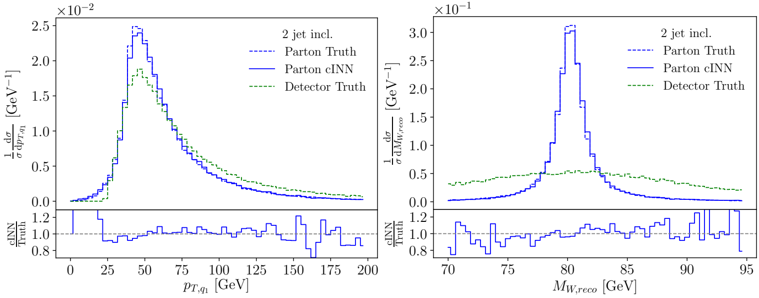
Inversion

Beyond forward simulation [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

– bijective transformation — physics mapped on latent space
  Jacobian tractable — normalizing flow [specifically: coupling layer]
  evaluation in both directions — INN [Ardizzone, Rother, Köthe]

– conditional GAN/INN: inverted events generated

– unfolding detector effects ... and jet radiation [$pp \to ZW \to (\ell\ell)\ (jj)$]



Inversion

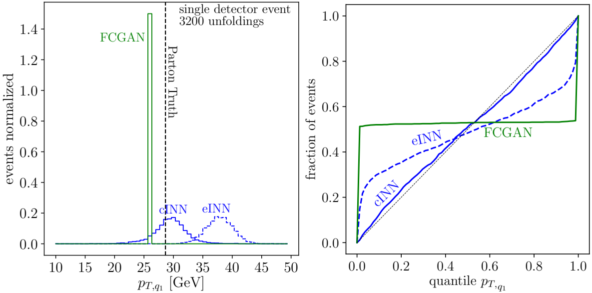
Beyond forward simulation [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

– bijective transformation — physics mapped on latent space
  Jacobian tractable — normalizing flow [specifically: coupling layer]
  evaluation in both directions — INN [Ardizzone, Rother, Köthe]

– conditional GAN/INN: inverted events generated

– unfolding detector effects ... and jet radiation [$pp \to ZW \to (\ell\ell)\ (jj)$]

⇒ parton-level pdf from single detector-level event



–

Inference

Fun things we can do with simulation

– where physics can tell ML people how to do it right

– Kyle Cranmer's talk

Heidelberg in virtual summer 2021