Machine Learning Ways to Improve LHC Analyses

Tilman Plehn

Universität Heidelberg

Berlin 1/2020

Useful theorists?

Fundamental understanding of LHC data

– LHC and dark matter data-driven, but never fundamental without theory

– just work with data and SM? [Jernej's talk]

1. simulation from first principles [Pythia, Sherpa]
2. interpretation frameworks [SMEFT, SUSY :)]
3. best use of the data [using 1, 2, our brains, and ML]

– 1991 visionaries: NN-based quark-gluon tagger [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.

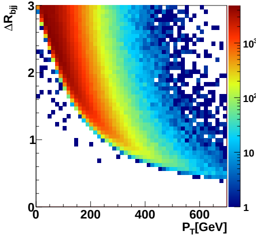
$\Rightarrow$ Lots of theory-related questions

Top tagging

Look what makes jets [Pre-LHC, jets were just annoying]

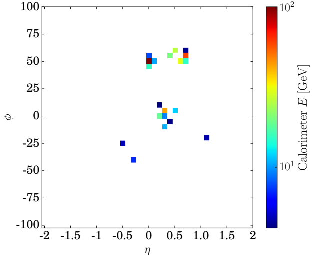
- top jets from $t \to b q \bar{q}'$ vs QCD jets
- top decays well-defined in theory
- labelled sample: semileptonic $t\bar{t}$ events

$\Rightarrow$ Fat jets as LHC physics playground

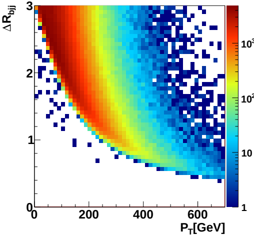


ML from low-level observables [Cogan etal, Oliveira, Nachman etal, Baldi, Whiteson etal (2014/15)]

- why intermediate high-level variables?
- calorimeter with all information



Top tagging



Look what makes jets [Pre-LHC, jets were just annoying]

– top jets from $t \to bq\bar{q}'$ vs QCD jets

– top decays well-defined in theory

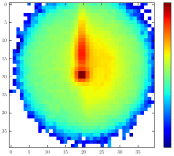
– labelled sample: semileptonic $t\bar{t}$ events

$\Rightarrow$ Fat jets as LHC physics playground

ML from low-level observables [Cogan etal, Oliveira, Nachman etal, Baldi, Whiteson etal (2014/15)]

– why intermediate high-level variables?

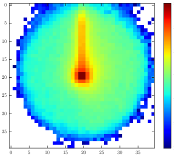
– calorimeter with all information

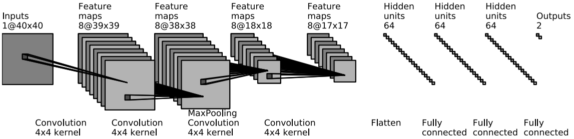


– tops from conv network [Kasieczka, TP, Russell, Schell; Macaluso, Shih]

– 40×40 bins through calorimeter resolution

– image recognition standard ML task





Theory inspiration

4-vector input — graph CNN [Butter, Kasieczka, TP, Russell; much better versions by now]

– physics objects from calorimeter and tracker

– distance measure known from e&m [alternatively: Erdmann, Rath, Rieger]

Inspired by QFT

– input 4-vectors $(k_{\mu,i})$

– jet algorithm $\longrightarrow$ combination layer

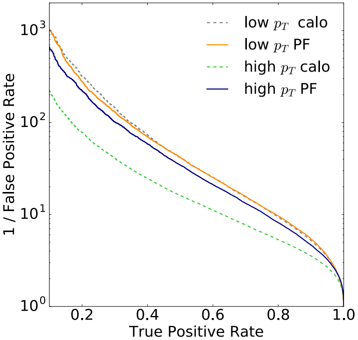
$$k_{\mu,i} \xrightarrow{\text{CoLa}} \widetilde{k}_{\mu,j} = k_{\mu,i}\; C_{ij}$$

– observables $\longrightarrow$ Lorentz layer

$$\tilde{k}_j \xrightarrow{\text{LoLa}} \hat{k}_j = \begin{pmatrix} m^2(\tilde{k}_j) \\ p_T(\tilde{k}_j) \\ \vdots \end{pmatrix}$$

$\Rightarrow$ Learn Minkowski metric

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



Jet classification done

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

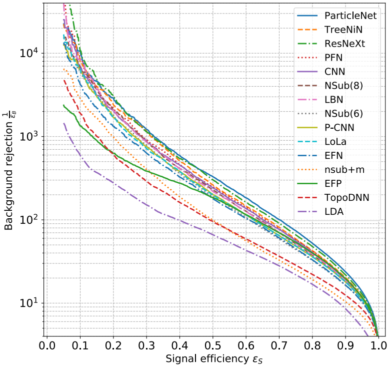
July 24, 2019

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.

Content

1 Introduction 3
2 Data set 4
3 Taggers 5
3.1 Imaged-based taggers 5
3.1.1 CNN 5
3.1.2 ResNeXt 6
3.2 4-Vector-based taggers 6
3.2.1 TopoDNN 6
3.2.2 Multi-Body N-Subjettiness 7
3.2.3 TreeNiN 8
3.2.4 P-CNN 8
3.2.5 ParticleNet 9
3.3 Theory-inspired taggers 9
3.3.1 Lorentz Boost Network 10
3.3.2 Lorentz Layer 11
3.3.3 Latent Dirichlet Allocation 11
3.3.4 Energy Flow Polynomials 12
3.3.5 Energy Flow Networks 13
3.3.6 Particle Flow Networks 14
4 Comparison 14
5 Conclusion 18
References 19



Beyond central values

Better Bayesian networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; cf Nachman]

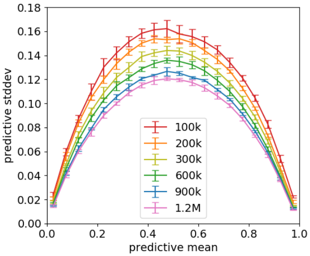
– neural network: deterministic functions
particle physics data: statistical
$\longrightarrow$ neural network output: statistical or wrong

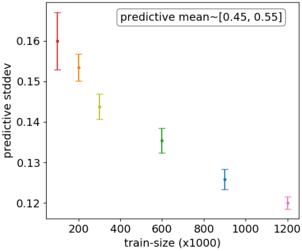
– classification: $(60\pm??)\%$ top?

– histogramming networks unrealistic
sampling network weights standard ['Bayesian' networks]

– formally: systematic approach to regularization and drop-out

– uncertainty from training statistics [parabola from closed interval output]

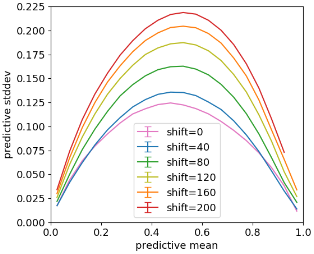


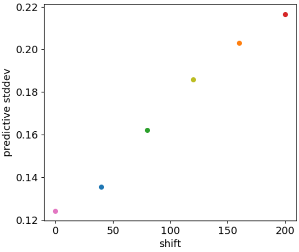


Beyond central values

Better Bayesian networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; cf Nachman]

- neural network: deterministic functions
  particle physics data: statistical
  $\longrightarrow$ neural network output: statistical or wrong
- classification: $(60\pm??)\%$ top?
- histogramming networks unrealistic
  sampling network weights standard ['Bayesian' networks]
- formally: systematic approach to regularization and drop-out
- uncertainty from training statistics [parabola from closed interval output]
- uncertainty from pile-up

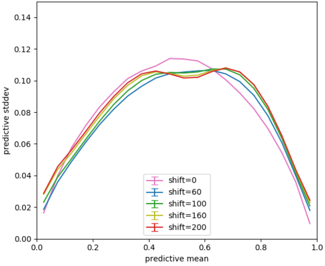




Beyond central values

Better Bayesian networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; cf Nachman]

- neural network: deterministic functions
  particle physics data: statistical
  $\longrightarrow$ neural network output: statistical or wrong
- classification: (60$\pm$??)% top?
- histogramming networks unrealistic
  sampling network weights standard ['Bayesian' networks]
- formally: systematic approach to regularization and drop-out
- uncertainty from training statistics [parabola from closed interval output]
- uncertainty from pile-up
- instability from pile-up



Beyond central values

Better Bayesian networks [Bollweg, Haussmann, Kasieczka, Luchmann, TP, Thompson; cf Nachman]

– neural network: deterministic functions
  particle physics data: statistical
  $\longrightarrow$ neural network output: statistical or wrong

– classification: $(60\pm??)$% top?

– histogramming networks unrealistic
  sampling network weights standard ['Bayesian' networks]

– formally: systematic approach to regularization and drop-out

– uncertainty from training statistics [parabola from closed interval output]

– uncertainty from pile-up

– instability from pile-up

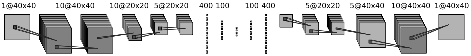
– tagger calibration part of the training

– performance just like usual taggers

  ....

– Lots of advantages, no cost

Learning from background



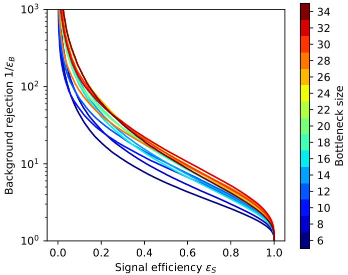
Hypothesis-free BSM searches [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih; Dillon etal]

– train on ‘background’, search for deviations

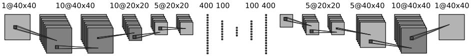
– established ML concept: autoencoder

– reconstruct typical QCD jet from data
reduce central weights, compress information
search for outliers

$\Rightarrow$ Making an okay tagger



Learning from background



Hypothesis-free BSM searches [Heimel, Kasieczka, TP, Thompson; Farina, Macari, Shih; Dillon etal]

– train on ‘background’, search for deviations

– established ML concept: autoencoder

– reconstruct typical QCD jet from data
reduce central weights, compress information
search for outliers

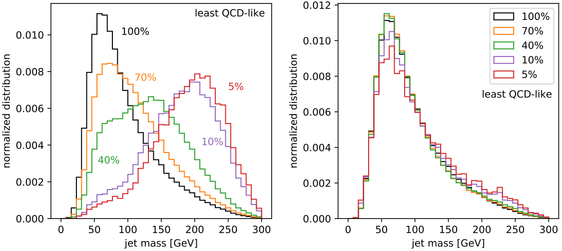
$\Rightarrow$ Making an okay tagger

De-correlate background shaping

– established ML concept: adversary [Cranmer, Shimmin; Spannowsky etal]

– atypical QCD jets typially with large jet mass
ignore jet mass from network training

$\Rightarrow$ Still open questions



Understanding networks

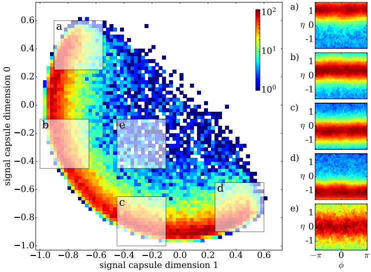
Capsule networks [Diefenbacher, Frost, Kasieczka, TP, Thompson]

– vector output instead of scalar classification [pooling vs stride]
– agreement by parallel vectors in feature space
– vector components for properties and geometry [eyes, nose, mouth]

Boosted tops from Z' resonance

– signal capsule for signal events
– two components distinctive through radius
– rotation remaining symmetric
– average event per region

signal identifying η_j
azimuthal angle insensitive



Understanding networks

Capsule networks [Diefenbacher, Frost, Kasieczka, TP, Thompson]

- vector output instead of scalar classification [pooling vs stride]
- agreement by parallel vectors in feature space
- vector components for properties and geometry [eyes, nose, mouth]

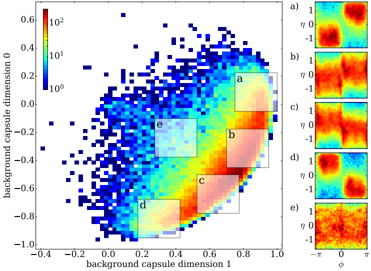
Boosted tops from Z' resonance

- signal capsule for signal events
- two components distinctive through radius
- rotation remaining symmetric
- average event per region

  signal identifying η_j
  azimuthal angle insensitive

  background identifying back-to-back

$\Rightarrow$ Useful?



The future

Machine learning is an amazing tool box...

...LHC physics really is big data

...imagine recognition is a starting point

...Bayesian networks with error bars

...capsule networks useful for visualization

Not even talked about GAN, reinforcement learning, and fun stuff...