Neural Networks in Top Jet Tagging

Priyotosh Bandyopadhyay, Abhishek Agarwal*
Indian Institute of Technology Hyderabad, Kandi, Sangareddy-502285, Telengana, India†

We consider Top Quark Production at LHC in pp collisions. Being the heaviest and third generation quark, Top Quark lies at the frontier of the current understanding of the Particle Physics and holds keys to the gates of Beyond Standard Model (BSM) Physics. Here we consider a novel way to identify Top Quarks at the LHC, using Deep Learning to classify Top Pair Production Events against QCD and other backgrounds.

I. INTRODUCTION

Many applications for Deep Learning Methods are currently employed by the CMS Collaboration at the LHC to identify heavy flavour quarks [1]. At the actual detector, many tools such as Multi Variate Analysis and Neural Networks are deployed and many results are often combined to reach a final conclusion about jet tagging and quark flavours. However so far, no single method has been employed to tag top flavoured jets. Many algorithms have been suggested [2] and quite efficient algorithmic results have been demonstrated in high energy jets.

Since top quark is very heavy, they are mostly created will decreasingly smaller momenta and consequently boost. High energy tops ($pT > 500GeV$) typically have orders of magnitude smaller cross section than tops with relatively small momenta ($pT > 20GeV$). This loss of cross section is a trade off while searching for highly boosted and collimated top jets, which are easier to identify. However with a more systematic and data driven approach to top jet identification although, it is possible to tag top jets without significant loss of cross section, as we demonstrate here.

II. METHOD

The outline of our new process is to first try to tag top jets as correctly as possible in simulated events such as Pythia[4]. This is not very complicated since in simulations, we can directly access flavour information at the parton level. Next we collect as many useful charaacteristics as we can about each and every jet and collect all the data in a list of jets format. Every row of this file (dataset) corresponds to 1 jet produced in the simulation along with its flavour (source). We then contruct observable variables for these jets and feed these observables along with source information to a Deep Neural Network (DNN). The DNN will identify all the features on its own. After training, the DNN will be able to tell us whether a given collection of observables is (most likely) a top jet or not.

Column Name	Observable	description
I	Yes	pT ordered index per Event
η	Yes	Pseudorapidity (round to 0.1)
ϕ	Yes	Azimuthal angle (round to 0.1)
pt	Yes	Transverse Momentum
$mult$	Yes	No. of particles in the jet
m	Yes	Invariant Mass
e	Yes	Relativistic Energy
$sibl$	Yes	No. of jets present in the event
q	Yes	Total Charge
Nch	Yes	No. of charged tracks
$girth$	Yes	$\frac{\sum_{i\ of\ J}(pT_i \Delta R_{i,J})}{pT_J}$
Xe	Yes	$\frac{max(E_i\ for\ i\ in\ Jet)}{E_J}$
$sumR$	Yes	$\sum_{i\ of\ J}(\Delta R_{i,J})$
src	No	Jet Flavour

TABLE I. Various properties collected for each jet

To implement our method, we went and simulated $pp \to t\bar{t}$ events. We used Pythia 8.2.4 built with FastJet 3.3.3 [5] for this purpose. For every event, we selected all particles that are stable, and visible in the detectors, but are not leptons and fed them into fastjet for jet clustering. FastJet then gives us a list of jet clusters for each event. In the next step we iterate through every jet and collect information about it. Table 1 lists all the variables collected for every jet.

The column src was calculated using a Mass Conctribution Method (MCM). For each Jet, $\{H_J\}$ is the set of the hardest parent of each constituent. We back track the parents of each i in $\{H_J\}$ until we find a parton from the hardest process. Since tops decay fully into b & W, for $pp \to t\bar{t}$ this means only three possibilities; we may find a t quark directly, or a W boson, or neither. We say $src_i = 2$ for a particle in $\{H_J\}$ if i comes from t quark directly (no W boson found while searching). Similarly $src_i = 1$ for those coming from W and $src_i = 0$ for those coming form neither. Then we define source for a jet as $src_J = k$ where $i \in \{H_J\}$ and:

$$\sum_{src_i=k} m_i = max(\sum_{src_i=0} m_i, \sum_{src_i=1} m_i, \sum_{src_i=2} m_i) \tag{1}$$

* bpriyo@phy.iith.ac.in, ep17btech11001@iith.ac.in
† https://iith.ac.in/~bpriyo

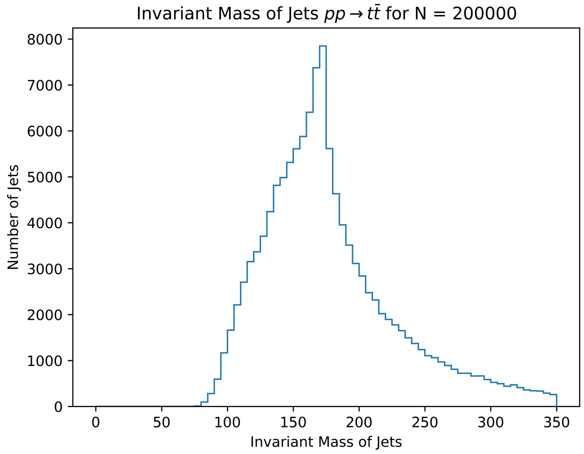

FIG. 1. $E_{CM} = 14\ TeV,\ |\eta_{jet}| \leq 2.5,\ pT_{jet} > 20\ GeV,\ R = 0.75$ with FastJet 3.3.3, Anti-k_t jet finding in Pythia 8.4.2

Process	$\sigma(fb)$	Signal
$pp \to t\bar{t}$	31.92	27281.7048
QCD Background	14110	271.6963577
$pp \to ZZ,\ ZW,\ WW$	1.92	0.7104
$pp \to H,\ HW,\ HZ$	30491	0.0
$pp \to Hf\bar{f},\ Ht\bar{t}$	0.000561	0.456654
$pp \to Wt\bar{t},\ Zt\bar{t}$	0.0	0.0
$pp \to ZZZ, WWW, WWZ, WZZ$	0.0	0.0

TABLE II. Background estimations: $pT_{min} \geq 1000\ GeV$

III. RECONSTRUCTION OF THE TOP QUARK

Given any event, we have a few jets with each of whom flavour has been associated. It is represented by one *src* column, which is 2 for b jet derivatives, 1 for W derivatives and 0 coming from neither. In order to be sure of our flavour assciation, we plot the invariant mass of jets. Instead of going into a complex process for composing these groups, we just do make simple combinations. First we make all ${}^{N}C_2$ pairs of jets in every event and plot their respective invariant mass. As expected, we see a peak at around 80 GeV, the peak for W boson mass. So we select two jets such that their invariant mass lies in the mass window (65, 95) GeV. Then we pick a third jet, whose sum of square of distances from the two jets is minimum, and we plot the invariant mass of all the three jets. This plot is shown in Figure 1. Given our simple approach to reconstruct top mass, our liberal mass windows of $(145, 205)\ GeV$ are compensating. We could try to make a very complicated method to catch much more top jets, and in turn tighten our mass window, but that is not our point. Our point is to be able to facilitate event classficiation.

Using this technique, we are able to count 31.05% events where at least one of the Top jets gave us the correct invariant mass. The striking part about this approach is that if we use the *src* column of from our simulation and reject all the jets whose *src* value is zero (background jets), we end up with the same number of jets within 2%. The other important bit is that since we have not used any pT cuts, we are able to make use of the croos cross section space.

However so far we cannot use this technique on experimental data. In order to do that we proceed to column wise tabulate our simulation data. We then feed this data including all the variables in Table 1 into a Deep Neural Network (DNN) and use it's learning to predict the *src* column for each jet. Instead of using 3 values for *src* we train our model to predict whether a given jet from any event had a zero or non zero *src* column, which is a binary classfiication problem, a well explored territory in DNNs. Since the jets with a $src_J = 0$ are background jets, we discard them and try to contruct the invariant mass of the top quark from the remaining jets.

The results are that the events for which at least one reconstructed jets' invariant mass lies in the range $(145, 205)$ GeV or tagging efficiency is 24.56%. The mistagging efficiencies are very low, well below 0.1%. This is a strong evidence that given any event which can be simulated, the DNN can learn the jet features of the simulation in order to assist us in potentially classifying real experimental events.

IV. BACKGROUND ESTIMATION

In this section we consider the potential background processes whihc may corrupt our data. Since the DNN learns jet features from the pure simulated events, but in real life will be tested on experimental data containing a wide variety of processes, we do a background estimation on the following processes in order to gauge the practical usability of our method. They were simulated with a min pT cut on the hard process as $pT > 1000GeV$. For comparison, the original top pair production is also included. Signal is calculated $= (\frac{n}{N})\sigma L$, where n is counted events (at least 2 jets must have non zero *src* value to be counted), N is total events and L is luminosity of choice $= 1000$. It is also to be noted that 1.6×10^7 QCD events were simulated due to limited computation power. The results are outlined in Table 2.

V. CONCLUSIONS

Neural Networks in High Energy Physics are not a new phenomena, people have been exploring many different ways to be able to leverage computational power in order to magnify signal background ratios. However we here propose a novel way to use today's sophesticated Monte Carlo simulators, and use Deep Networks trained on simulation data in order to classify experimental data.

As we can see, the Top Quark Pair Production event gives a signal about two order of magnitude larger than most backgrounds. These backgrounds which complicate our life a great deal in experimental data, force us to move to higher pT events, restricting our cross section. However if we employ features from simulation data and use them to classify events in experimental data, our backgrounds could be reduced substantially without any significant loss in the signal. At the data driven stage, we correctly classify 85.45% of the Top Pair Production events, while being able to reconstruct top mass in 31.05% of the events. However at the testing level, after putting in a naive neural network, we correctly classify 68.54% of the $t\bar{t}$ events, and reconstruct the invariant mass of Top Quark in 24.56% events. The mistagging from background events rises from 2×10^{-4}% to about 3×10^{-2}%. A high varability in mistagging is observed, with the Neural Network structure and efficiency making a large difference in mistagging rates, athough all such observed mistagging rates were well below 0.1%. The significance is calculated as $\frac{singal}{\sqrt{signal+background}}$ which comes out to be around 163. Such a high confidence tells use that we might have a lot of space for exploring such new techniques and presents a stong evidence for putting this technique to test.

In this report, we looked at a very simply jet recombination scheme along with the simplest flavour association parameters one could think of, resulting in negligible mistagging and roughly a quarter of the events classified correctly without any cross section cuts. This is convincing at least as a proof of concept, while actual application and analysis on real experimental data will be needed to assess the future of such techniques.

[1] Sirunyan, A.M. and Tumasyan, A. and Adam, W. and Ambrogi, F. and Asilar, E. and Bergauer, T. and Brandstetter, J. and Brondolin, E. and Dragicevic, M. and Erö, J. and et al., J. of Instr., (2018), [10.1088/1748-0221/13/05/p05011 [ISSN:1748-0221]]

[2] D. E. Kaplan, Rehermann, Keith, Schwartz, Matthew D. and Tweedie, Brock, Phys. Rev. L., (2008), [10.1103/physrevlett.101.142001]

[3] A. Belyaev, N. D. Christensen and A. Pukhov, Comput. Phys. Commun. **184** (2013) 1729 [arXiv:1207.6082 [hep-ph]].

[4] T. Sjostrand, L. Lonnblad and S. Mrenna, [arXiv:hep-ph/0108264].

[5] M. Cacciari, G. P. Salam and G. Soyez, Eur. Phys. J. C **72** (2012) 1896 [arXiv:1111.6097 [hep-ph]].