

Deep Learning Techniques for Top-Quark Reconstruction

Kiarash Naderi

Sharif University of Technology

CERN Summer Student Programme

August 2017

Abstract

Top quarks are unique probes of the standard model (SM) predictions and have the potential to be a window for physics beyond the SM (BSM). Top quarks decay to a Wb pair, and the W can decay in leptons or jets. In a top pair event, assigning jets to their correct source is a challenge. In this study, I studied different methods for improving top reconstruction. The main motivation was to use Deep Learning Techniques in order to enhance the precision of top reconstruction.

1 Introduction

Several reasons make top quarks notable particles for investigating SM and probing BSM. Top quarks are the heaviest particles in the SM. Their mass is 173.34 ± 0.27 GeV [1]. Due to that fact the lifetime of top quark is short. Which opens a window for studying bare quarks [2]. The invariant mass of top-antitop quarks ($m_{t\bar{t}}$) is an interesting quantity due to the fact that is sensitive to important physical points such as the top quark mass; QCD predictions for PDFs, the strong coupling (α_s), and bound-states; resonant production through new physics e.g. $Z' \rightarrow t\bar{t}$; and electroweak corrections for the the particles Z , photon, and Higgs.

Besides the physical motivations for investigating top quarks, there are several practical challenges. For instance, top quarks need to be reconstructed from their decay products and neutrinos are not easily detectable. Most importantly, assigning jets to their origin is a challenge. In this project, starting from basic reconstruction methods such as δR and χ^2 , as explained in the followings, finally I arrived at using a Deep Neural Network (DNN) to improve the reconstruction of the $t\bar{t}$ system.

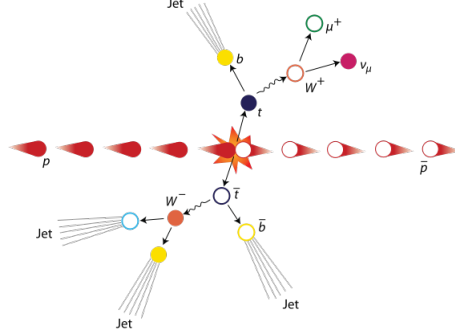


Figure 1: Top decay

2 Basic reconstruction methods

2.1 The δR approach

A top quark decays to a Wb pair. With some probability jets from W are close enough in the phase space due to a slight boost of the W boson in the laboratory. This idea leads us to the δR reconstruction method: to reconstruct the W boson from the light jets, select the two closest ones.

2.2 The χ^2 approach

The χ^2 minimization is an approach that is helpful for testing the goodness-of-fit [4]. In particular, according to Figure 1, among the all different combinations of jets, I select the combination that minimizes the χ^2 expression:

$$\chi^2 = \frac{(m_W - m_{LJ_1+LJ_2})^2}{r_{W_1}^2} + \frac{(m_W - m_{\nu+l})^2}{r_{W_2}^2} + \frac{(m_t - m_{LJ_1+LJ_2+BJ_1})^2}{r_{t_1}^2} + \frac{(m_t - m_{\nu+l+BJ_2})^2}{r_{t_2}^2} \quad (1)$$

It selects the combination that in the best way belongs to the top quarks decay. LJ and BJ respectively stand for light jets and b-jets. m_W and m_t are the masses of the W boson and the top quark. Finally, r is the resolution of the detector for a specific decay that is known from the simulation. Table 1 shows the values of the parameters in the Equation 1 that I used in GeV.

r_{W_1}	r_{W_2}	r_{t_1}	r_{t_2}	m_W	m_t
7.0	7.0	10.0	10.0	80.4	172.4

Table 1: Numerical values for the χ^2 approach in GeV

2.3 Comparison between the two approaches

Figure 3 depicts the resolution (i.e. $\frac{m_{t\bar{t}}}{m_{gen_t}} - 1$ where m_{gen_t} is the "truth" mass of the $t\bar{t}$ from the simulation) for the two approaches. It is apparent

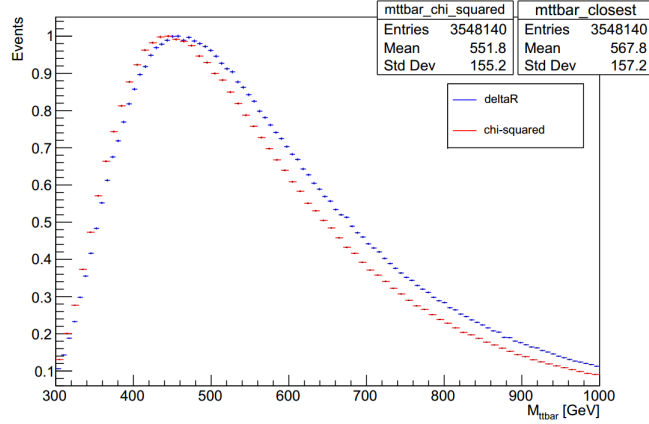


Figure 2: The reconstructed mass of the $t\bar{t}$ system.

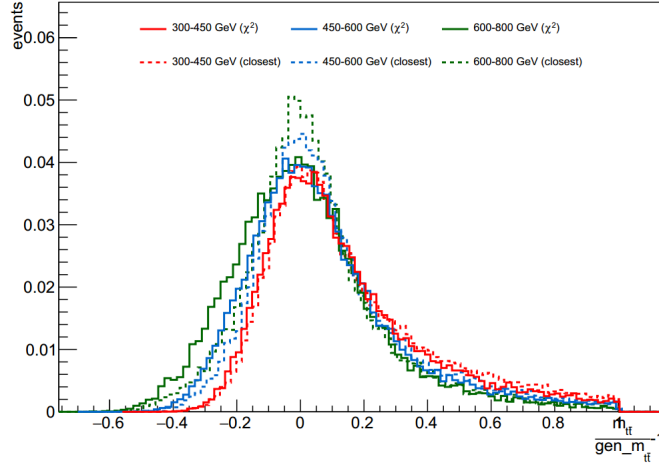


Figure 3: The plot of $\frac{m_{t\bar{t}}}{m_{gen_t}} - 1$ for different ranges of m_{gen_t} .

from Figure 4 the χ^2 approach performs significantly better for high jets multiplicity.

3 Deep Neural Network

A neural network consists of several *nodes* and *layers*. Each layer consists of nodes that are the places where the computations happen. Generally speaking, each node is fed with inputs that are effective through a weight. The weight of the connections between the nodes are trained using about 1 M simulated $t\bar{t}$ events with the full detector simulation. The sample is weighted to a flat $m_{t\bar{t}}$ distribution to avoid a bias from the simulated distribution. Then the output goes to other nodes on the next layer [5].

A Deep Neural Network (DNN) is a neural network with more than one hid-

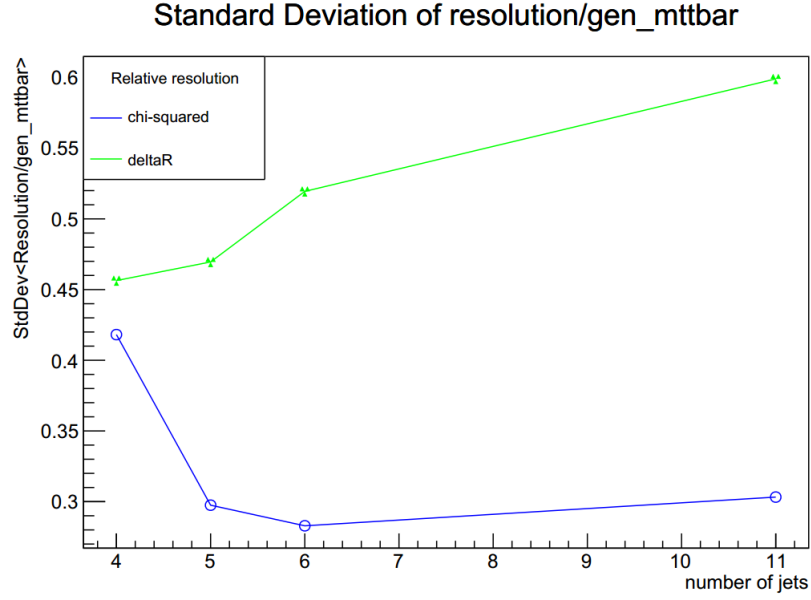


Figure 4: Standard deviation of the resolution over the jet multiplicity.

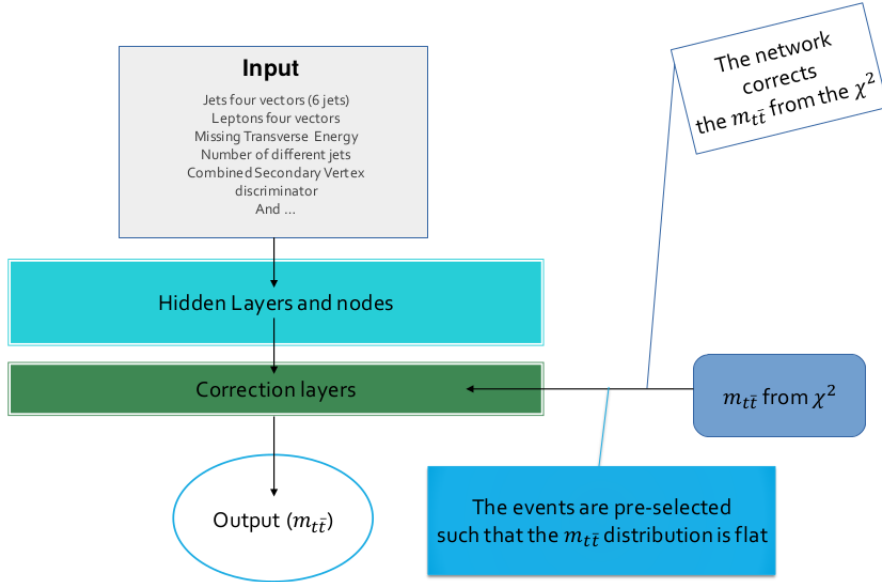


Figure 5: The structure of a typical DNN that I used.

Techniques	Benefits
Adding more layers	To let the network grasp more non-linear and sophisticated dependencies.
Increasing epochs of trainings and reducing the learning rate	Network can reach the truth value with minor steps rather jumping around the truth value.
Adding dropout layers	Prevents the network from overfitting.

Table 3: The machine learning techniques that I used.

den layer. Theoretically, every function can be estimated up to an arbitrary precision with a neural network. However, in practice it very depends on the structure of the problem.

In my case, I used a DNN to correct the $m_{t\bar{t}}$ of the χ^2 approach. Figure 5 shows the generic structure that I used.

3.1 DNN results

Table 2 shows the inputs to the network and Table 3 consists of the studied techniques. Following is the structure of the best model I have studied:

- 4 layers, 128 nodes
- All the 4 layers shrinks into a node and then they are added to the $m_{t\bar{t}}$ from the χ^2 as a correction.
- I used GaussianDropout technique (see Ref. [6]) that smears the inputs.

Figures 6 and 7 depict the resolution of this DNN in compared to the χ^2 approach. Table 4 shows the mean value and the standard deviation for the four different approaches.

Jets four-vectors	Leptons four-vectors	MET
Number of different jets	CSV b-jet discriminator	$m_{t\bar{t}}$ from the χ^2

Table 2: The inputs to the network.

	δR	χ^2	DNN	[3]
Mean value	10%	7%	4 %	1%
Standard Deviation	25.1%	25.0%	17.2%	15.9%

Table 4: Comparison of the three different approaches

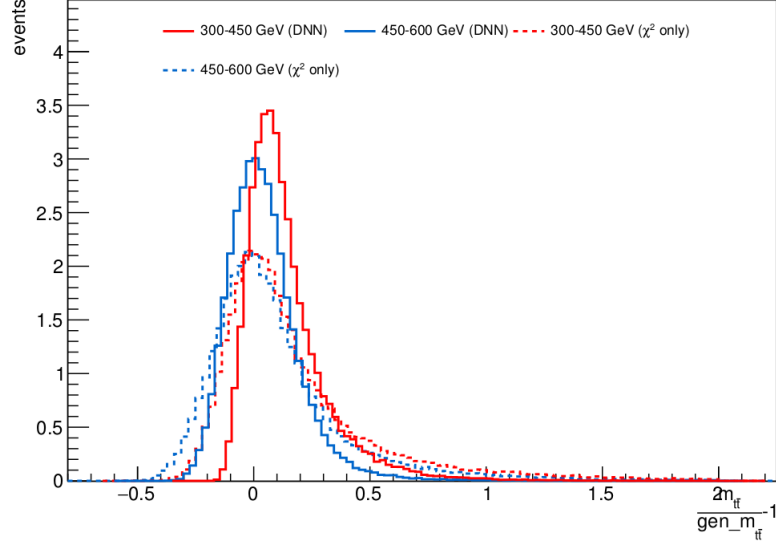


Figure 6: The resolution of the DNN and the χ^2 approaches for $m_{gen_t} \leq 600$ GeV.

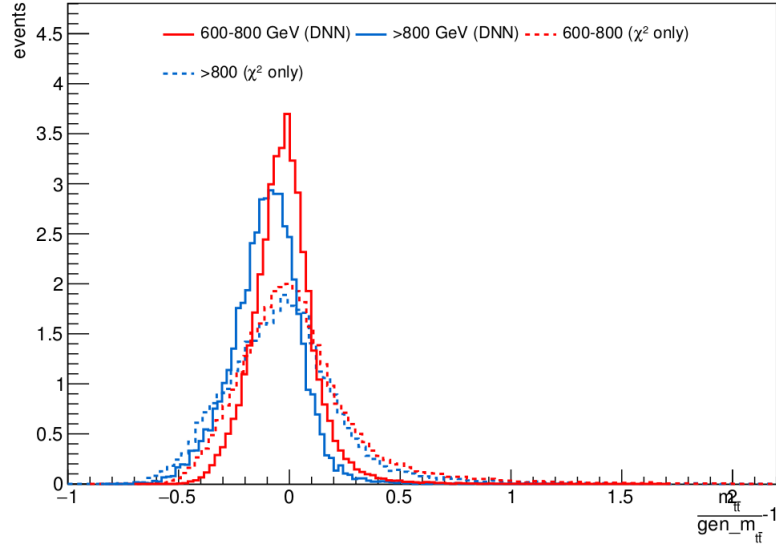


Figure 7: The resolution of the DNN and the χ^2 approaches for $m_{gen_t} > 600$ GeV.

4 Conclusion

The main goal of this project was to assess if the reconstruction of the $m_{t\bar{t}}$ variable can be improved by using machine learning techniques. My studies have shown that is possible to improve over a baseline χ^2 -based method significantly (by almost 30%). With respect to the method employed in Ref. [3], that is the best algorithm so far in the CMS TOP group, the results are shown to be competitive, differing by $\approx 2\%$. My result shows that neural networks can be potentially useful for improving the LHC measurements in the Run 2 and beyond. There is still plenty of room for optimization.

5 Acknowledgments

I would like to thank Otto Hindrichs for kindly providing the $m_{t\bar{t}}$ resolution for comparison. My deepest appreciation goes to Jan Kieseler and Pedro Ferreira da Silva for kindly supervising me during the project.

References

- [1] The ATLAS, CDF, CMS and DØ Collaborations. *First combination of Tevatron and LHC measurements of the top-quark mass*. arXiv:1403.4427.
- [2] Michele Gallinaro. *Top quark physics: A tool for discoveries*. LIP Lisbon. Journal of Physics: Conference Series 447 (2013) 012012.
- [3] The CMS Collaboration. *Measurement of differential cross sections for top quark pair production and associated jets using the lepton+jets final state in proton-proton collisions at 13 TeV*. Phys. Rev. D 95 (2017).
- [4] W. R. Leo. *Techniques for Nuclear and Particle Physics Experiments: A How to Approach*. ISBN: 9783540572800.
- [5] Michael Nielsen. *Neural Networks and Deep Learning*. Determination Press 2015.
- [6] Srivastava, Nitish and Hinton, Geoffrey E and Krizhevsky, Alex and Sutskever, Ilya and Salakhutdinov, Ruslan. *Dropout: a simple way to prevent neural networks from overfitting..* Journal of Machine Learning Research, 2016-11-17.