IL NUOVO CIMENTO **45 C** (2022) 110
DOI 10.1393/ncc/i2022-22110-0

Communications: SIF Congress 2021

Regression Deep Neural Networks for top-quark-pair resonance searches in the dilepton channel

G. Guerrieri(1)(2)(*)

(1) *INFN, Sezione di Trieste - Trieste, Italy*
(2) *Università degli Studi di Udine - Udine, Italy*

received 28 January 2022

Summary. — Several Beyond Standard Model (BSM) theories predict the existence of new massive particles decaying to pairs of top quarks, $t\bar{t}$. In this concept work, the key observable for such resonance searches, the top-pair system invariant mass, $m_{t\bar{t}}$, is reconstructed by training a deep neural network on a sample of simulated $t\bar{t}$ events. A regression task is then performed on both $t\bar{t}$ and Z' signal events, using $m_{t\bar{t}}$ as output parameter. The comparison between this machine learning approach and more traditional system reconstruction techniques highlights a tangible improvement in the ability to correctly reconstruct and resolve a TeV-scale $t\bar{t}$ resonance peak.

1. – Introduction

Massive particles decaying into top quark-antiquark pair ($t\bar{t}$) [1] are predicted by many theoretical models, which are introduced to provide explanations to the various open questions raised by the current formulation of the Standard Model of Particle Physics (SM). Due to its large mass, the top-quark decays before hadronizing, with a lifetime of about $\mathcal{O}(10^{-25})$ s. Therefore, the top-quark has to be reconstructed from its decay products, and since it decays 99.8% of the times to a W boson and b-quark [2], the final-state topologies in $t\bar{t}$ production are strongly dependent on the decay modes of the W bosons. The study of $t\bar{t}$ resonances constitutes a cornerstone of the physics programme of the ATLAS experiment [3] at the CERN Large Hadron Collider (LHC) [4], in which BSM studies on $t\bar{t}$ are typically performed in the semileptonic and dileptonic channels. These conventional $t\bar{t}$-resonance searches rely on the comparison of the data yield with the expected SM-background, taking into account the reconstructed $t\bar{t}$ invariant mass distribution, $m_{t\bar{t}}$ in each bin. New physics signals would be seen as a localized excess

(*) E-mail: giovanni.guerrieri@cern.ch

of data events with respect to the SM-background expectation, or, in case of strong interference effects, as a deficit, or even as a non-trivial peak-dip structure [5]. As a matter of fact, the dilepton channel allows to achieve a strong sensitivity to the $t\bar{t}$ spin correlation through variables as simple as the azimuthal angle difference between the two charged leptons $\Delta\phi_{\ell\ell}$, due to the large spin analyzing power of leptons, close to unity. The main drawback, due to the presence of two undetected neutrinos, is that the quantity $m_{t\bar{t}}$ is not directly accessible. To precisely reconstruct $m_{t\bar{t}}$, a regressive Deep Neural Network (DNN) is employed; DNNs belong to a subset of Machine Learning algorithms which are meant to improve automatically their performance through experience, without being explicitly programmed to perform a specific task and with few or no assumptions on the underlying physical processes.

2. – Simulated event samples and model architecture

The datasets used for the DNN training and testing consist of SM events generated through a full Monte Carlo (MC) simulation of the ATLAS detector performed with GEANT4 [6]. The simulated samples emulate p-p collisions at $\sqrt{s} = 13\,\mathrm{TeV}$ and the LHC Run 2 beam conditions, and represent $t\bar{t}$ events decaying in the dilepton channel, generated at the NLO-QCD accuracy by the Powheg program [7], interfaced to Pythia8 for simulating the parton shower and hadronization steps [8]. Events are required to pass an initial base selection, demanding at least two reconstructed hadronic jets and exactly two high transverse-momentum leptons (electrons or muons) with same or opposite flavour, and opposite charge.

Each of the 40 million events that composes the training dataset is characterised by a label, representing the MC truth value of the $m_{t\bar{t}}$ variable, and by a collection of 22 features which correspond to the measured four momenta of the two charged leptons and of up to three jets in the event, as well as the available information on the missing energy in the transverse plane. The four momenta are expressed in terms of the energy (E), the transverse momentum (p_T), the azimuthal angle (ϕ) and the pseudorapidity (η), in addition to the missing transverse energy (E_T^{mis}) and the associated angle (ϕ^{mis}). If less than three jets are produced, the missing measurements are filled with the 0 value.

Training events are selected in order to form a dataset in which the label $m_{t\bar{t}}$ is uniformly distributed, *i.e.*, in which the energy bins are set to reach a maximum number of counts. The dataset has sufficient statistics to guarantee a $m_{t\bar{t}}$ distribution approximately uniform up to $2\,\mathrm{TeV}$; this characteristic enables the possibility to exploit the trained DNN to test BSM physics, in particular the production of a $t\bar{t}$ pair originating from the hypothetical Z' decay [9].

The model chosen for the analysis consists of a deep feed-forward neural network, structured as a 22-nodes input layer, three hidden layers, each one with 100 nodes and a single-node output, estimating the value of $m_{t\bar{t}}$ for each event. The model architecture is summarized in table I. After each layer, a rectified linear activation function is used. The Mean Absolute Error (MAE) is taken as loss function; the Adam optimiser [10] is used to minimise the loss function with respect to the network parameters. The DNN training comprises a few tens of cycles over the full set of event samples (epochs). Each epoch is batched into subsamples of 50 events, with a learning rate set to 0.0001. The architecture of the DNN, implemented with the Keras [11] software package, and the training parameters were chosen from several trials to ensure reproducible and optimal performance.

TABLE I. – *Schematic summary of the model chosen for data analysis.*

Layer	Connection	Output Shape	Parameters
Hidden (1)	dense	(none, 100)	2300
Hidden (2)	dense	(none, 100)	10100
Hidden (3)	dense	(none, 100)	10100
Output	dense	(none, 1)	101

3. – Validation with MC data

The DNN is deployed to reconstruct $m_{t\bar{t}}$ of both SM and BSM simulated events; The results are compared with those given by two other more traditional methods, which are briefly described below.

3˙1. *llbb*. – This method consists in the calculation of the partial invariant mass m_{llbb} of the top-antitop pair, given by

$$m_{llbb} = \sqrt{\left(p^{l^+} + p^{l^-} + p^{b_1} + p^{b_2}\right)_\mu \left(p^{l^+} + p^{l^-} + p^{b_1} + p^{b_2}\right)^\mu}, \tag{1}$$

where $p^{l^\pm}$ and $p^{b_{1,2}}$ are the four-momentum vectors for electron or muon and b-tagged jets, respectively. This method is considerably straightforward, thus often adopted; the obvious drawback is that the contribution of the two undetected neutrinos to $m_{t\bar{t}}$ is just neglected.

3˙2. *Neutrino weighting*. – The neutrino weighting technique [12] relies on scanning over a set of possible values for the two neutrino pseudorapidities, solving a system of equations for each possible pair of values, and then checking the compatibility of the found solutions with the measured value of E_T^{mis}. The solution with the largest weight is selected. However, the performance of this technique suffers from the non-negligible fraction of events where no suitable solution is found, especially for high values of $m_{t\bar{t}}$.

The Pearson correlation coefficient ρ between true and predicted values of $m_{t\bar{t}}$ is considered to precisely estimate the accuracy of the various methods. Even if still in a preliminary stage, the comparison between the correlation coefficients for each technique, shown in table II, suggests a better efficiency of the DNN in reconstructing $m_{t\bar{t}}$, both in the case of SM and BSM events.

4. – Conclusions

In this work, a feed-forward Deep Neural Network was implemented and it proved to be a useful tool for reconstructing the invariant mass of top quark pairs decaying in the dilepton channel. Comparing this innovative method with some of the most commonly used traditional analysis techniques, such as the usage of the invariant mass of the visible decay products only ("llbb") as a proxy for $m_{t\bar{t}}$, as well as a partially analytic solution for neutrino reconstruction ("neutrino weighting"), it was possible to show significant

TABLE II. – *Summary of the values of the Pearson correlation coefficient measured for each of the considered $m_{t\bar{t}}$ reconstruction method.*

	DNN	llbb	Neutrino weighting
ρ^{SM}	0.82	0.59	0.61
ρ^{BSM}	0.43	0.28	0.23

improvements in mass estimation in both SM and BSM scenarios. Several upgrades to the analysis are undergoing development: considering additional high-level features such as b-tagging, exploiting different architectures for the neural network, and providing a multi-dimensional output prediction, *i.e.*, adding to $m_{t\bar{t}}$ other significant variables predictions, such as a spin-correlation–sensitive angular observable.

This novel regression method is planned to be adopted in searches for a heavy Higgs boson decaying to $t\bar{t}$ pairs [13], as well as in a possible Z' search.

* * *

I gratefully thank Daniele Tilotta and the Gruppo Collegato INFN Trieste - Università di Udine for the feasibility study on the subjects discussed in this work [14].

REFERENCES

[1] THE ATLAS COLLABORATION, *Eur. Phys. J. C*, **80** (2020) 528.
[2] ZYLA P. *et al.*, *Prog. Theor. Exp. Phys.*, **08** (2020) 083C01.
[3] THE ATLAS COLLABORATION, *J. Instrum.*, **3** (2008) S08003.
[4] EVANS L. and BRYANT P., *J. Instrum.*, **3** (2008) S08001.
[5] FREDERIX R. and MALTONI F., *J. High Energy Phys.*, **01** (2009) 047.
[6] AGOSTINELLI S. *et al.*, *Nucl. Instrum. Methods Phys. Res. Sect. A*, **506** (2003) 250.
[7] ALIOLI S. *et al.*, *J. High Energy Phys.*, **06** (2010) 043.
[8] SJÖSTRAND T. *et al.*, *Comput. Phys. Commun.*, **191** (2015) 159.
[9] DEL AGUILA F., *Acta Phys. Pol. B*, **25** (1994) 1317.
[10] KINGMA D. P. and BA J., *Adam: A Method for Stochastic Optimization* (2017) `https://doi.org/10.48550/arXiv.1412.6980`.
[11] CHOLLET F. *et al.*, *Keras* (2015) `https://github.com/fchollet/keras`.
[12] THE D0 COLLABORATION, *Phys. Lett. B*, **752** (2016) 18.
[13] THE CMS COLLABORATION, *J. High Energy Phys.*, **04** (2020) 171.
[14] TILOTTA D., `https://cds.cern.ch/record/2763283` (Dec. 2020).