

Machine learning approaches for parameter reweighting in Monte-Carlo samples of top quark production in CMS

Valentina Guglielmi^{1,*}

*Deutsches Elektronen-Synchrotron (DESY),
Notkestrasse 85, Hamburg, Germany*

E-mail: valentina.guglielmi@desy.de

In high-energy particle physics, complex Monte Carlo (MC) simulations are needed to compare theory predictions to measurable quantities. Many and large MC samples are needed to be generated to take into account all the systematics. Therefore, the MC statistics (and hence the MC modeling uncertainties) become a limiting factor for most measurements. Moreover, the significant computational cost of these programs becomes a bottleneck in most physics analyses. Therefore, it is extremely important to find a way to reduce the MC samples generated to decrease the MC statistical uncertainties and lower the computational cost. In these proceedings, we evaluate an approach called Deep neural network using Classification for Tuning and Reweighting (DCTR). DCTR is a method based on a Deep Neural Network (DNN) to reweight simulations to different models or model parameters and fit simulations, using the full kinematic information in the event. This reweighting methodology avoids the need for simulating the detector response multiple times by incorporating the relevant variations in a single sample. In this way, the MC statistical uncertainties and the computational cost are both reduced. Moreover, unlike the standard reweighting, in which the ratio in bins of two histograms at truth level is performed, multidimensional and unbinned information can be used as inputs to the DNN. In addition, DCTR can perform tasks that are not possible with other current existing methods, such as continuous reweighting as a function of any MC parameter, simultaneous reweighting of more MC parameters and tuning MC simulations to the data. We test the method on MC simulations of top quark pair production, which we reweight to different SM parameter values and to different QCD models.

*41st International Conference on High Energy physics - ICHEP2022
6-13 July, 2022
Bologna, Italy*

¹For the CMS Collaboration.

*Speaker

1. Introduction

In high-energy particle physics, complex Monte-Carlo (MC) simulations are needed to compare theory predictions to the measurable quantities, i.e. data. Many and large MC samples are needed to be generated to take into account all the systematics. For this reason, MC modeling uncertainties are a limiting factor for most measurements. For example, in a recent measurement of the top quark-antiquark ($t\bar{t}$) pair production cross section in pp collisions at a centre of mass energy $\sqrt{s} = 13$ TeV, where a simultaneous fit of the cross section of the $t\bar{t}$ system and of the MC mass of the quark top m_t^{MC} was performed [1], the main contribution to the m_t^{MC} was given by the MC statistics of the samples used to estimate the systematics. Furthermore, the generation of many MC samples significantly increases the computational cost of these programs requiring to simulate the detector response multiple times. A possible solution is to reweight the MC sample. In this case, only the MC sample with the parameter nominal values is generated, while the variations are obtained by reweighting the nominal sample. Employing this strategy, only a sample is generated reducing MC statistics (hence MC modeling uncertainties) and lowering the computational cost.

Unlike the standard reweighting, in which the ratio in bins of two histograms at truth level is performed, Deep neural network using Classification for Tuning and Reweighting (DCTR) is presented in these proceedings. DCTR is an approach based on a Deep Neural Network (DNN) to reweight simulations to different models or model parameters using the full kinematic and flavor information in the event [2]. While the standard reweighting is sensitive to the binning chosen and the dimension of the inputs could be 1, or at maximum 2, due to the increasing difficulty of the method with the number of dimensions, there are no restrictions on the size of the input feature space nor on the number of interpolated parameters using a DNN. The first feature permits improving the reweighting precision using all the kinematic and flavor information of the event as inputs. The second feature has as a consequence that simultaneous reweighting of more MC parameters can be performed, taking into account the correlations between the parameters. Moreover, in addition to the advantages of the reweighting using a DNN, DCTR permits to perform several tasks not possible with the other methods. First of all, DCTR is not limited to applying a discrete reweighting, but permits to perform a continuous reweighting as a function of any MC parameter. In this way, a fast and more precise estimation of the systematic uncertainties is possible since any parameter value can be extrapolated. Secondly, in addition to reweighting, the method can be also exploited to directly tune MC simulations to the data. In this case, a classifier is used to construct the loss function used to optimize the simulation parameters.

2. Deep neural networks using Classification for Tuning and Reweighting

The default MC sample of top pair production in CMS is generated using the Heavy Quark Process (HVQ) [4] of the event generator Powheg [5][6]. In Powheg, the resummation of the next-to-leading-order radiation is regulated by the h_{damp} variable, which enters the damping parameter D as in Eq. 1:

$$D = \frac{h_{\text{damp}}^2}{p_T^2 + h_{\text{damp}}^2} \quad (1)$$

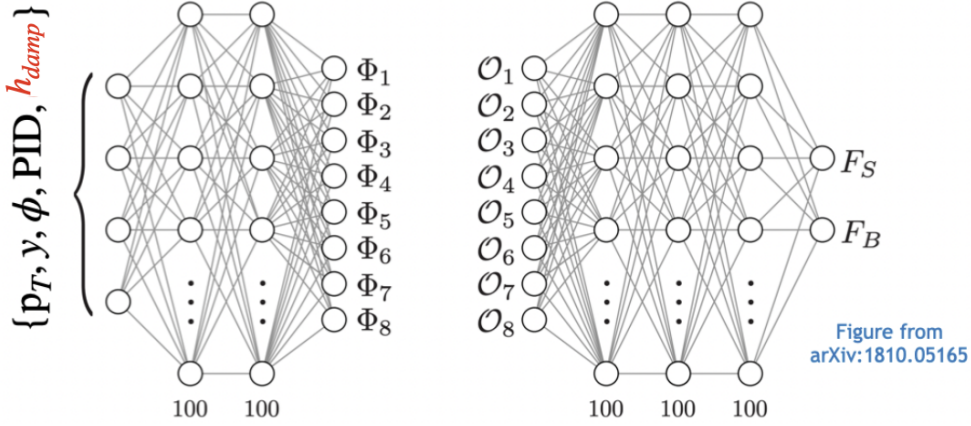


Figure 1: The PFN architecture used in the DCTR method is depicted. It parametrizes the per-particle mapping Φ (on the left) and the function F (on the right), shown for the case of a latent space of dimension $l = 8$. The latent observable is $O_a = \sum_i \Phi_n(p_T, y_i, \phi_i, m_i, pdgid_i)$ [3].

where p_T is the transverse momentum of the particle and h_{damp} a parameter defined as $h_{\text{damp}} = h \times m_t$, where m_t is the mass of the quark top and h is a real number. This parameter affects the kinematic information of the entire event, such as the transverse momentum and the pseudorapidity of the $t\bar{t}$ system ($p_T(t\bar{t}), \eta(t\bar{t})$). Since the parameter h_{damp} is not physical, an arbitrary value must be chosen in the simulation and varied to calculate the associated systematic uncertainty. As the parameter cannot be reweighted internally by MC generators, and h_{damp} variations are important in many top quark precision studies, this parameter is especially well suited to be implemented in the DCTR method. Therefore, variations of h_{damp} in $t\bar{t}$ system were studied to test how well the method works.

3. Results and Conclusions

The DCTR method was tested by performing the training on samples of four different values of h_{damp} , respectively equal to $(0.5 \times m_t, 1 \times m_t, 2 \times m_t, 4 \times m_t)$, where m_t corresponds to the default value of CMS (172.5 GeV). For each value of h_{damp} , two million events were generated. The parton-level information is passed as inputs to the DNN. The quadrimomentum and the particle PDGID ($p_T, y, \phi, m, \text{PDGID}$) of the $t\bar{t}$ system and of the additional quark or gluon are presented to the DNN for training, plus the reweighting parameter h_{damp} . The Particle Flow Network (PFN) [6], which is the composition of two DNNs (Φ and F) is used. The DNN architecture can be seen in Fig. 1.

The impact of the parameter h_{damp} on the p_T of the $t\bar{t}$ system can be seen on the left plot in Fig. 2. To check the performance of the reweighting, the observable p_T of the $t\bar{t}$ system was employed. The weights obtained in the training were applied to reweight a distribution with h_{damp} value equal to $4 \times m_t$ to the nominal one generated with h_{damp} value equal to $1 \times m_t$. The results

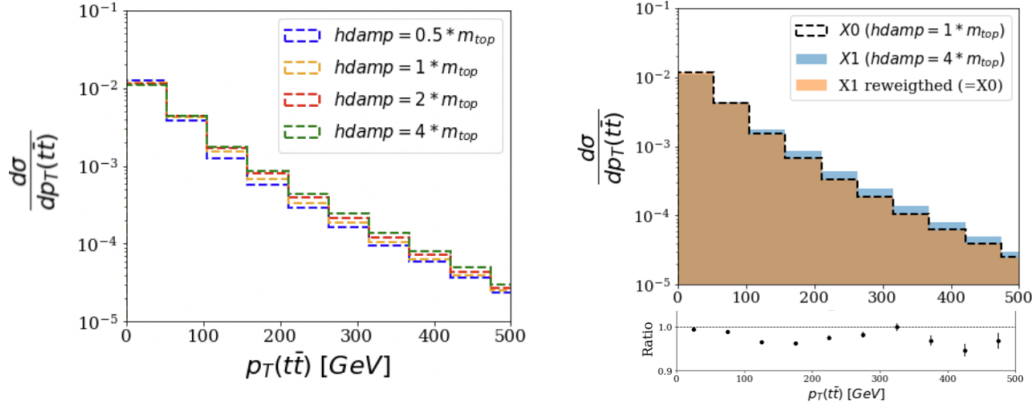


Figure 2: Left: The differential cross section as a function of the p_T of the $t\bar{t}$ system for Powheg HVQ samples generated with four different values of the h_{damp} parameter. Right: The differential cross section in p_T of the $t\bar{t}$ system as a function of the transverse momentum of the $t\bar{t}$ system for a sample generated with $h_{\text{damp}} = 1 \times m_t$ in black and $h_{\text{damp}} = 4 \times m_t$ in blue. The orange distribution is the second sample (blue) reweighted to the first one (black).

are shown on the right plot in Fig. 2. The precision of the reweighting is quantified by the ratio plot between the reweighted sample (orange) and the original one (black). The original sample and the reweighted one agree within an uncertainty of 5%. The next step would be to integrate the DCTR method into the CMS software to be used in all future top analyses.

References

- [1] CMS collaboration, *Measurement of the $t\bar{t}$ production cross section, the top quark mass, and the strong coupling constant using dilepton events in pp collisions at $\sqrt{s} = 13$ TeV*, *EPJC* **79** (2019) 368 [1812.10505].
- [2] A. Andreassen and B. Nachman, *Neural Networks for Full Phase-space Reweighting and Parameter Tuning*, *PRD* **101** (2020) 091901 [1907.08209].
- [3] P.T. Komiske, E.M. Metodiev and J. Thaler, *Energy Flow Networks: Deep Sets for Particle Jets*, *JHEP* **01** (2019) 121 [1810.05165].
- [4] S. Frixione, P. Nason and G. Ridolfi, *A Positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, *JHEP* **09** (2007) 126 [0707.3088].
- [5] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with Parton Shower simulations: the POWHEG method*, *JHEP* **11** (2007) 070 [0709.2092].
- [6] S. Alioli, P. Nason, C. Oleari and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, *JHEP* **06** (2010) 043 [1002.2581].