PROCEEDINGS OF SCIENCE

Towards quantum-inspired Machine Learning on high-energy physics data at LHCb

Davide Zuliani,[a,b,*] **Timo Felser,**[a,b,c] **Marco Trenti,**[a] **Lorenzo Sestini,**[b] **Alessio Gianelle,**[b] **Donatella Lucchesi**[a,b] **and Simone Montangero**[a,b]

[a] *University of Padova, Padova, Italy*
[b] *INFN Sezione di Padova, Padova, Italy*
[c] *Universität des Saarlandes, Saarbrücken, Germany*

E-mail: davide.zuliani@cern.ch

The analysis of data produced in proton-proton collisions at the Large Hadron Collider (LHC) is very challenging and it will require a huge amount of resources when High Luminosity LHC will be operational. Recently, Machine Learning methods have been employed to tackle this task, with high efficiency but low interpretability. In this study [1] a possible application of a quantum-inspired algorithm based on tree tensor networks to study simulated data at LHCb is shown, in order to properly classify $b\bar{b}$ di-jet events and to interpret the classification result.

40th International Conference on High Energy physics - ICHEP2020
July 28 - August 6, 2020
Prague, Czech Republic (virtual meeting)

*Speaker

PoS(ICHEP2020)931

1. Introduction

At the Large Hadron Collider (LHC) protons collide at a center-of-mass energy of 13 TeV, producing a variety of particles. In particular, quarks produced after the collision cannot exist as free particles, and they manifest themselves as bound states (hadrons) or as narrow streams of particles produced by hadronization (jets). At LHCb [2], a forward spectrometer designed to study the forward region of proton-proton collisions, it is interesting to study jets coming from heavy-flavor quarks (namely b and c quarks), and therefore it is necessary to perform a good separation between different flavors. The technique used to achieve such a task is called jet flavor tagging. In this study [1], the identification of jets produced by b- and $\bar{b}$-quarks is taken into consideration, since it is fundamental to measure the $b\bar{b}$ charge asymmetry [3], which could be sensitive to New Physics processes [4]. At LHCb, b-jets are tagged using several methods, one of these relies on a single-particle tagging algorithm called muon tagging: the charge of the muon with highest transverse momentum p_T inside the jet is used to tag the b-quark, since the muon charge is completely correlated with the b-quark charge. In our study we would like to show that a quantum-inspired method based on tensor networks can classify b- and $\bar{b}$-jets. We also compare its performances with the standard LHCb tagging method and state-of-the-art deep neural network.

2. Tree Tensor Network

Tensor Networks (TNs) are a mathematical tool developed to investigate quantum many-body systems on classical computers [5]. They rely on an efficient representation of a quantum many-body wavefunction $|\psi\rangle$ in a compact form, by approximating a high-order tensor by a set of low-order tensors that are contracted in a particular underlying geometry: here we consider a Tree Tensor Network (TTN). The accuracy of the approximation is controlled by an auxiliary parameter called *bond-dimension* χ, controlling the amount of information captured within the ansatz. Recently it has been proved that TNs can be applied to solve supervised Machine Learning (ML) problems [6]. Exploiting their original development for simulating quantum systems, TNs allow computing typical quantities such as correlations and entanglement entropy, giving more insights into the learning process.

3. Methodology

Monte Carlo samples from LHCb Open Data [7] are considered for proton-proton collisions at center-of-mass energy of 13 TeV, producing $b\bar{b}$ di-jet events. Some kinematic cuts are considered to select proper events: jets are required to have transverse momentum $p_T > 20$ GeV and pseudorapidity η in the range $2.2 < \eta < 4.2$. A dataset of about $700k$ jets is then split into two datasets: 60% of the jets are used in the training process while the remaining are used as a test set. Inside each jet several types of particles are selected: $\mu^{\pm}$, $e^{\pm}$, $\pi^{\pm}$, $K^{\pm}$ and $p/\bar{p}$ and for each type the particle with the highest p_T is selected; particle types are selected by exploiting LHCb excellent performances in Particle IDentification (PID). Finally, for each of the selected particles three variables are selected: the charge q, the transverse momentum relative to the jet axis p_T^{rel} and the distance between the particle and the jet axis in the (η, ϕ) space ΔR (where ϕ is the azimuth

angle and $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$). Finally, the jet charge Q, defined as the sum of the particle charges weighted for p_T^{rel}, for a total of 16 variables. These variables are used as inputs for two inclusive ML classifiers, a Deep Neural Network (DNN) and a TTN, therefore exploiting the jet substructure. For each event prediction, both classifiers give as output a probability $\mathcal{P}_b$ for a jet to be generated by a b- or $\bar{b}$-quark. The classification probability $\mathcal{P}_b$ is interpreted in the following way: for values $\mathcal{P}_b > 0.5$ ($\mathcal{P}_b < 0.5$) a jet is classified to be produced by a b-quark ($\bar{b}$-quark), with an increasing confidence going to $\mathcal{P}_b = 1$ ($\mathcal{P}_b = 0$). The figure of merit for this kind of study is the tagging power ε_{tag}, defined as

$$\varepsilon_{tag} = \varepsilon_{eff}\,(2a-1)^2 \tag{1}$$

where ε_{eff} is the efficiency (fraction of tagged jets) and a is the accuracy (fraction of corrected tagged jets). The tagging power represents the effective fraction of jets contributing to the statistical uncertainty in the $b\bar{b}$ charge asymmetry measurement. A threshold on the high uncertainty region of $\mathcal{P}_b$ is applied in order to maximize the tagging power for each considered algorithm. Finally performance from both ML algorithms are compared with the standard muon tagging approach used so far at LHCb.

4. Results

The tagging power ε_{tag} as a function of the jet p_T for the two classifiers and the muon tagging approach is shown in Fig.1: both ML algorithms outperform the standard muon tagging approach by a factor $\sim$ 10 and they perform similarly, showing better performances for lower p_T. The scatter plot of TTN and DNN output is also presented, highlighting a correlation between the two classifiers.

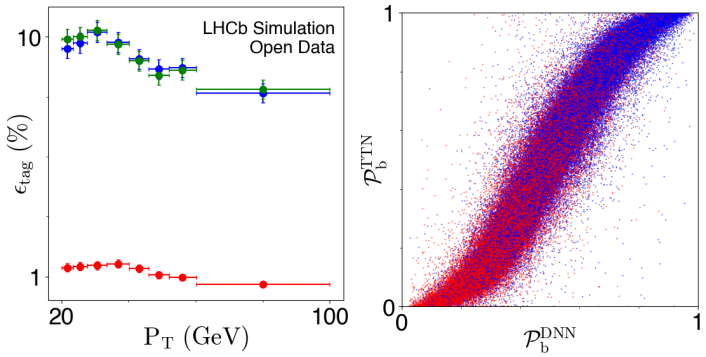

Figure 1: Comparison of the DNN and TNN analysis: (left) tagging power for the DNN (green), TTN (blue) and the muon tagging (red), (right) scatter plot of DNN and TTN predictions with $b(\bar{b})$-quarks in blue (red). Images taken from [1].

Further insights are obtained by looking at the probability distributions for the two classifiers. In Fig.2 the probability distributions are presented and it is evident that, despite both classifiers have the same performances, the probability distributions are different: for the DNN (left) we see a Gaussian-like shape which gives less correct predictions with high confidences and fewer wrong predictions with high confidences; for the TTN (right) we have a flat distribution including more predictions, correct and incorrect, with higher confidence. An interesting feature for the TTN is the

presence of peaks at $\mathcal{P}_b = 0, 1$ which are coming from jets containing a muon, since the presence of a muon inside the jet is a well-defined predictor for a jet generated by a b-quark. The DNN lacks this confident predictions.

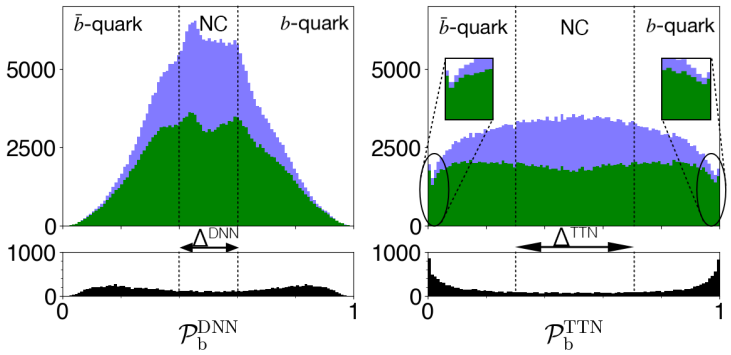

Figure 2: Probability distribution for the DNN (left) and the TTN (right). The correctly classified events (green) are shown in the total distribution (light blue). Below, in black all samples where a muon was detected in the jet. Images taken from [1].

The TTN framework enables us to capture correlations and entanglement within the classifier, in this way it is possible to identify the most important features typically used by the classifier in the classification process. Measuring correlations between features allow us to understand the relations between features: if two features are highly correlated or anti-correlated the information of one of the features can be neglected and gained back by using the other feature; on the contrary, if there is no correlation between features it means that both features may provide fundamentally different information for the classification. In Fig.3 correlations for the 16 variables in the b-quark classification are shown. The correlations analysis is not sufficient to claim that the information coming from those variables is important for the classification, along with the relative information of the feature we need to know the actual magnitude, it is, therefore, necessary to measure the entanglement entropy S of each feature, as shown in Fig.3. In the ML context, the entanglement entropy S can be interpreted as the quantity of information contained in a set of features affecting the classification. In this way it is possible to discard those non-informative features for the classification, introducing therefore a new model with fewer features and complexity.

The *Quantum Information Post-learning feature Selection* (QuIPS) algorithm exploits these considerations, allowing us to rank and therefore to reduce the number of features for the classification; in this way, we select the best 8 features to classify the events: charge, momenta and distance of the muon, charge, momenta, and distance of the kaon, charge of the pion, the total charge of the jet. In Fig.4 we compare the tagging power for the model with the best 8 features (B_8 model) with the one composed of the worst 8 (W_8 model), the complete model, and the muon tagging approach: it is evident that despite halving the number of features, the results between the complete model and B_8 are still comparable, while the W_8 model is performing worse than the standard tagging approach.

Finally, we study the possibility to reduce the prediction time, since short prediction times are required to achieve real-time event selection: during Run 2 data-taking LHCb collected data approximately every 1μs. To do so we exploit an interesting feature of the TTN: by adjusting the bond

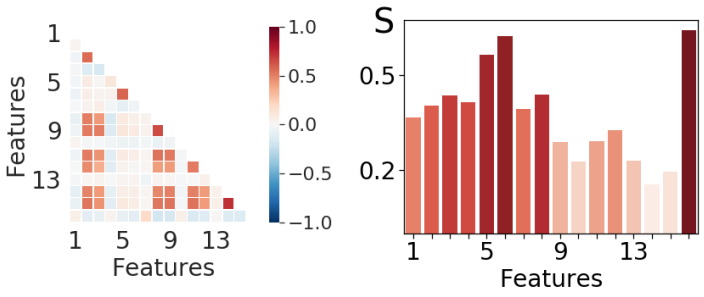

Figure 3: (left) Correlations between the 16 input features (blue for anti-correlated, white for uncorrelated, red for correlated). The numbers indicate q, p_T^{rel}, ΔR of the muon (1-3), kaon (4-6), pion (7-9), electron (10-12), proton (13-15) and the jet charge Q (16). (right) Entropy of each feature as measure for the information provided for the classification. Images taken from [1].

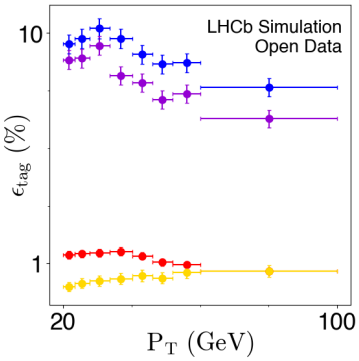

Figure 4: TTN tagging power for learning on all features (blue, complete model), the best 8 (B_8 model) proposed by QuIPS exploiting correlations and entanglement measurements (magenta), the worst 8 (W_8 model, yellow) and the muon tagging (red). Image taken from [1].

dimension χ it is possible to target a specific prediction time while keeping the prediction accuracy reasonably high. This can be done after the training procedure, therefore without relearning a new model, as would be the case for a neural network. This is achieved via the *Quantum Information Adaptive Network Optimization*, which is combined with the QuIPS algorithm to reduce the information used by the TTN in an optimal way by balancing the prediction time and the accuracy. In Fig.5 results are shown for the complete and the B_8 model, with different bond dimensions χ: by compressing down the complete model to $\chi = 5$ the overall accuracy does not change significantly, while improving the prediction time from 345μs to 37μs. Applying the same idea to the B_8 model we reduce the prediction time to 19μs, which is compatible with the current real-time classification rate. This study has been performed only on one CPU processor core, therefore there could be some improvements by performing the tensor contractions on GPUs.

5. Conclusions

In this study we analyzed the possibility to apply TTN to supervised ML problems such as classifying b- and $\bar{b}$-jets at LHCb. We compared the TTN performance to a DNN and we pointed out

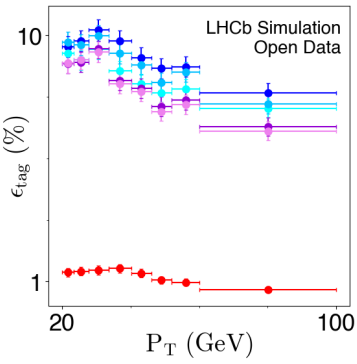

Figure 5: TTN tagging power for decreasing bond dimension truncated after training: The complete model (blue shades for $\chi = 100$, $\chi = 50$, $\chi = 5$), for using the QuIPS best 8 features only (B_8 model, violet shades for $\chi = 16$, $\chi = 5$), and the muon tagging (red). Image taken from [1].

that, despite the two classifiers use information in different ways, their performance results similar. Both ML classifiers outperform the standard muon tagging approach by one order of magnitude. By further analyzing the TTN framework we exploited the possibility to study relations between features, namely by measuring correlation and entanglement entropy: in this way we proved that is possible to select the best features and still get good performances. We also performed a truncation of the network after the training step to reduce the classification time, obtaining performances comparable to the current real-time classification rate. Future studies will focus on studying b- and c-jets separation and performing real-time applications.

References

[1] Felser, T., Trenti, M., Sestini, L. *et al.* Quantum-inspired machine learning on high-energy physics data. npj Quantum Inf 7, 111 (2021). https://doi.org/10.1038/s41534-021-00443-w

[2] A. A. Alves Jr. *et al.* (LHCb collaboration), *The LHCb detector at LHC*, JINST 3 (2008) S08005.

[3] R. Aaij *et al.* (LHCb collaboration), *First measurement of the charge asymmetry in beauty-quark pair production at a hadron collider*, Phys. Rev. Lett. 113, 082003 (2014)

[4] K. Fraser and M.D. Schwartz, JHEP 10, 093 (2018).

[5] S. Montangero, *Introduction to Tensor Network Methods*, Springer International Publishing, 2018.

[6] Stoudenmire, E. and Schwab, D. J, *Supervised learning with tensor networks*, In (eds Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I. and Garnett, R.) Advances in Neural Information Processing Systems 29, 4799–4807. http://papers.nips.cc/paper/6211-supervised-learning-with-tensor-networks.pdf (Curran Associates, Inc., 2016).

[7] R. Aaij *et al.*, (LHCb collaboration), *LHCb open data website*, http://opendata.cern.ch/docs/about-lhcb (2020).