Institute for Research in Fundamental Sciences

مقاله نامه ی بیست و ششمین کنفرانس بهاره فیزیک- ۲۲-۲۳ خرداد ماه ۱۳۹۸

شماره (۱۶)

Performance of fully-connected neural networks for top tagging

Javad Ebadi[1,2] and Daruosh Haji Raissi[2]

[1]School of Physics, Institute for Research in Fundamental Sciences (IPM)
[2]School of Particles and Accelerators, Institute for Research in Fundamental Sciences (IPM)

Abstract

We compare performance of several fully-connected neural networks for hadronically decaying top quarks. Although fully-connected neural networks have the simplest architecture, they still have a good performance for identifying boosted top quarks. We show that the performance of a simple fully-connected deep neural network could reach accuracy of about 89%.

1 Introduction

The standard model (SM) of particles physics does not have an answer for problems such as dark matter or matter anti-matter asymmetry. The solution for these problems lies beyond the SM (BSM). There are BSMs which can be tested at the LHC. However, in some cases the search can be challenging and intricate data analysis methods are needed.

Tagging a top quark is essential for data analysis at the LHC. The common way to tag top quark is to look for its well separated decay products, i.e. W boson and b quark. However, this method does not work for boosted top quarks, because the decay products are collimated. Boosted top quarks at the LHC can be produced in BSM models. For instance, in Randull-Sundrum models, the lightest Kaluza-Klein gluon can decay to a pair of boosted top quarks [1]. Thus, in order to search for new physics, we need new techniques for tagging top quark.

There are plenty of methods for top tagging based on QCD. The Johns Hopkins University top tagger (JHUTopTagger) is a well known QCD based top tagger [2]. In recent years, machine learning methods, especially deep neural network has been used for top tagging. In machine learning method, the machine could learn features to distinguish top quarks jets from quark or gluon jets without knowing anything about QCD [3].

In this work, we compare performance of fully-connected neural networks for top tagging.

2 Model and data

We get the data from the DESY cloud associated to [4] paper. The data contains 2M events, 1.2M for training, 400k for validation and 400k for testing. The data are either hadronic tops (signal) or QCD di-jets (background) for the 14 TeV LHC. The pile-up effects is not considered in the analysis. The fat jets are cluster using anti-k_T algorithm with $R = 1.5$.

Data normalization is very important in machine learning algorithms. The training process can be much faster with an appropriate data normalization. In this work, we have normalized each feature as

$$p_{j,\mu}^{(m)} \rightarrow \frac{p_{j,\mu}^{(m)}}{\sum_j E_j^{(m)}} \tag{1}$$

where m, j and μ are indices for event, particle and 4-vector, respectively.

We have used keras [5] with Theano[6] backend to build neural networks. Adam optimizer [7] is used in training of neural networks. In hidden layers, the ReLU functions is used for activation and in output layer, sigmoid function is used as activation. The loss function is binary crossentropy which is a common loss function for binary classifications problems.

3 Result

Receiver Operating Characteristic (ROC) curves are plotted in Fig. 1 and area under the curve (AUC) is shown in Tab. 1. In addition, number of parameters in models, number of nodes in hidden layers and accuracy are shown in Tab. 1. The models are identified with names such as H*FCDNN*. The first number after H letter determines number of hidden layers, and the last number determines number of nodes in each hidden layer of the network. The first model which

Table 1: Information of fully-connected neural networks.

model	**parameters**	**hidden layers nodes**	**accuracy**	**AUC**
Logistic regression	801	-	0.703	0.772
H1FCDNN32	25665	[32]	0.793	0.870
H1FCDNN128	102657	[128]	0.822	0.897
H2FCDNN32	26721	[32,32]	0.884	0.939
H3FCDNN32	27777	[32,32,32]	0.892	0.947

is a logistic regression (LR) model (a neural network without any hidden layers) has accuracy of 0.703. This result may not be good, but since the number of parameters are very low, the model is robust and safe from overfitting. In addition, feature extraction can increase the performance of LR model.

The models with hidden layers have significantly better classification power than just plain LR model. In addition, we see that a network with many hidden layers (H3FCDNN32) can outperform a network with just one hidden layers but with a large number of parameters (H1FCDNN128).

The shortcoming of deep neural network is that as the number of layers increases the training becomes slower, because updating early layers parameters by optimizer is very slow. We have used a neural style transform like approach to solve this problem. In particular, to train a network with 3 hidden layers, first we trained a network with 2 hidden layers and then before starting to train the large network, we set the weights of the first two layers of the large network equal to the small network weights.

Among neural network architectures the best accuracy till now is 93.7% [4]. As we have shown here, a neural network with 3 hidden layers and 2777 parameters can reach accuracy of 89.2% which is a promising result considering the simplicity of the model.

4 Conclusion

We showed that as a neural network gets deeper the performance of the network increases. However, training of a deep neural network is very time consuming. To overcome this shortcoming we proposed a technical way to train a deep network.

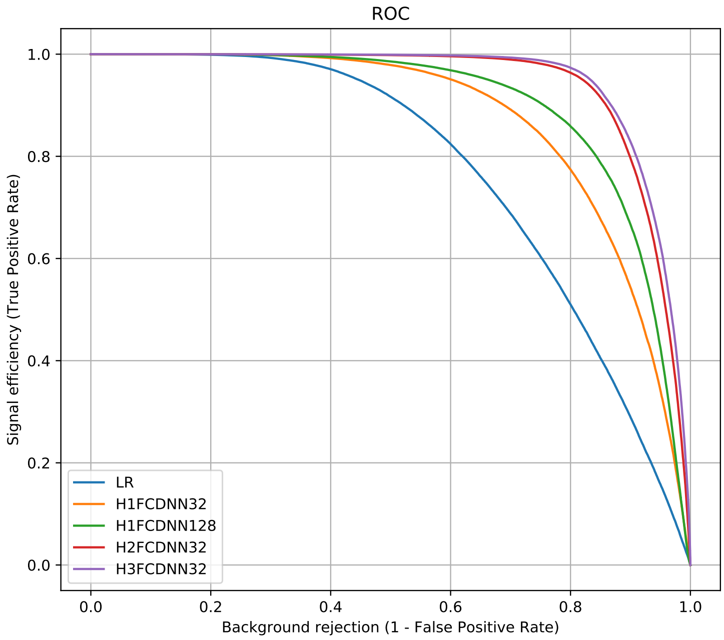

Figure 1: The ROC curves for classifiers.

Acknowledgement

The authors would like to thank M. Mohammadi Najafabadi for reading the draft and useful comments.

References

[1] B. Lillie, L. Randall, and L.-T. Wang, "The Bulk RS KK-gluon at the LHC," *JHEP* **09** (2007) 074, `arXiv:hep-ph/0701166 [hep-ph]`.

[2] D. E. Kaplan, K. Rehermann, M. D. Schwartz, and B. Tweedie, "Top Tagging: A Method for Identifying Boosted Hadronically Decaying Top Quarks," *Phys. Rev. Lett.* **101** (2008) 142001, `arXiv:0806.0848 [hep-ph]`.

[3] A. Butter *et al.*, "The Machine Learning Landscape of Top Taggers," `arXiv:1902.09914 [hep-ph]`.

[4] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, "Deep-learned Top Tagging with a Lorentz Layer," *SciPost Phys.* **5** no. 3, (2018) 028, `arXiv:1707.08966 [hep-ph]`.

[5] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[6] R. Al-Rfou, G. Alain, A. Almahairi, C. Angermueller, D. Bahdanau, N. Ballas, F. Bastien, J. Bayer, A. Belikov, A. Belopolsky, *et al.*, "Theano: A python framework for fast computation of mathematical expressions," `arXiv:cs.SC/1605.02688 [cs.SC]`.

[7] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," `arXiv:cs.LG/1412.6980 [cs.LG]`.