

Development of Machine Learning based muon trigger algorithms for the Phase2 upgrade of the CMS detector

T. Diotallevi*

University of Bologna, Italy

E-mail: tommaso.diotallevi@studio.unibo.it

D. Bonacorsi

University of Bologna and INFN, Italy

C. Battilana

University of Bologna and INFN, Italy

L. Guiducci

University of Bologna and INFN, Italy

After the high-luminosity upgrade of the LHC, the muon chambers of CMS Barrel must cope with an increase in the number of interactions per bunch crossing. Therefore, new algorithmic techniques for data acquisition and processing will be necessary in preparation for such a high pile-up environment. Using Machine Learning as a technique to tackle this problem, this paper focuses on the production of models - with data obtained through Monte Carlo simulations - capable of predicting the transverse momentum of muons crossing the CMS Barrel muon chambers, comparing them with the transverse momentum assigned by the current CMS Level-1 trigger system.

Sixth Annual Conference on Large Hadron Collider Physics (LHCP2018)

4-9 June 2018

Bologna, Italy

*Speaker.

1. Introduction

The current CMS Level-1 muon trigger system [1] performed smoothly in Large Hadron Collider (LHC) Run 1 and, after the "Phase 1" upgrade in 2015, in Run 2 as well. However with the advent of the High-Luminosity phase of the LHC (HL-LHC), the instantaneous luminosity will increase up to $7.5 \cdot 10^{34} \text{ cm}^{-2}\text{s}^{-1}$, approximately five times larger than the limit reached during the present LHC run. This corresponds to an average pile-up of up to 200 events per crossing in the interaction region of the detector and an integrated luminosity of up to 4000 fb^{-1} over 10 years of data taking. Such level of machine performance will allow to extend searches for new physics and to perform stringent tests of the Standard Model, such as precision measurements of the Higgs Boson couplings. For this reason, the CMS detector and its trigger system will need to undergo a substantial upgrade, called "Phase2 Upgrade", affecting all subdetectors: tracking, electromagnetic and hadronic calorimeters, muon detectors, trigger and readout systems. The overall Software and Computing systems will need to be completely revisited, too: given the higher complexity of the event reconstruction, using simulations for the upgrade phase, the computing power required is about 65-200 times worse than in the current run (Run-2).

The complexity and the time span of this challenge, together with the recent ramp-up in the evolution curve of selected advanced computing techniques - like Machine Learning (ML) and Deep Learning (DL) approaches - invites to explore some of them and to implement actual prototypes that test and verify their feasibility and possible adoption. The work done so far towards ML/DL-based barrel muon trigger algorithms for the Phase2 upgrade of the CMS detector will be presented and discussed.

2. The barrel muon Level-1 trigger system

The Level-1 trigger is designed to take a fast accept/reject decision (every bunch crossing or BX), in a pipeline mode using custom developed programmable hardware. The Level-1 Muon trigger is designed to reconstruct muon position and transverse momentum (p_T) and assign particle origin in terms of BX. The functional relations between the components are shown in Figure 1 (left). The electronics devoted to muon tracking follows a geographical partitioning: separate systems process local information produced in the barrel, endcap and overlap regions, corresponding to $|\eta| < 0.8$, $0.8 < |\eta| < 1.2$ and $1.2 < |\eta| < 2.4$, respectively, to identify muons and measure their coordinates and momentum (Figure 1 right).

The barrel region, which this paper focuses on, is composed by two different families of gaseous detectors: *Drift Tubes* (DT) and *Resistive Plate Chambers* (RPC). The information is then collected and synchronized by the TwinMux system, which is in charge of the merging of the two, thus delivering improved trigger elements (*superprimitives*) to the *Barrel Muon Track Finder* (BMTF) [2].

2.1 Barrel Muon Track Finder

The muon barrel architecture groups the barrel muon detectors in 12 wedges. Each wedge has five sectors and each sector four DT stations and six RPC planes. The front-end electronics generate the muon primitives and send the data to the TwinMux. Then it fans out the superprimitives to the

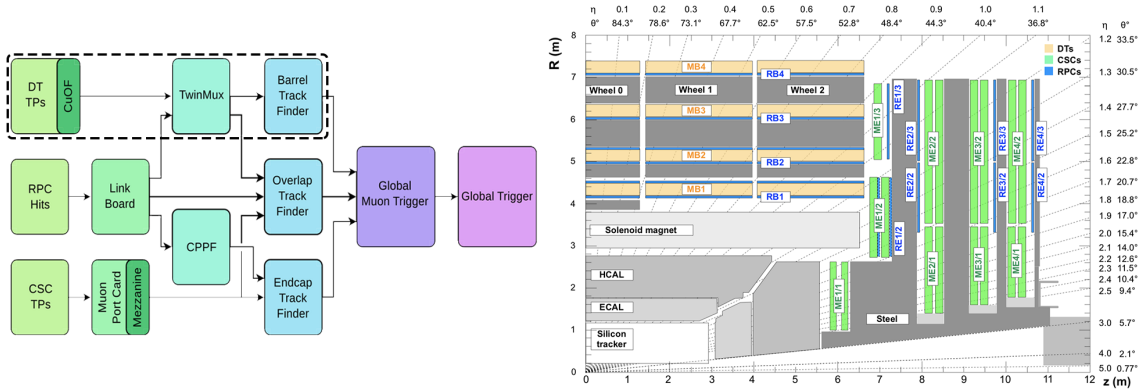


Figure 1: On the left, the CMS Level-1 Muon Trigger data flow (the dashed circle highlights the barrel region of the detector). On the right, a longitudinal view of the CMS muon region is given.

Barrel Muon Track Finder (BMTF).

The BMTF trigger receives the muon superprimitives from the barrel area of CMS ($|\eta| < 0.8$) and uses this information to reconstruct muon tracks and calculate the physical parameters: p_T , the ϕ angle, η , and the quality of the candidate. The algorithm combines the track stubs, bending angle and quality bits from the innermost stations; then it links the segment pairs to the full track and assigns it with a quality word related to the number and type of stations involved. Finally, using a Look-Up Table (LUT), the p_T physical parameter is assigned to the algorithm output.

The value of p_T assigned after the output produced by the Barrel Muon Track Finder will be used as a comparison for the p_T 'predicted' by the Machine Learning models.

3. Artificial Neural Network with Keras and TensorFlow

An artificial neural network (ANN) is a learning method inspired by the biological neural connections that constitute a human brain, specifically designed to non-linear learning problems [3]. Currently, ANNs are one of the main tools used in Machine Learning. Each input feature is represented as a neuron, creating an input layer: the output(s) are then computed by a weighted sum of the inputs, similarly to other learning algorithms. The main difference between Neural Networks and other algorithms relies in the addition of other layers of neurons between them, called hidden layers.

The Application Programming Interface (API) environment used for the Machine Learning model - deeply specialized in Neural Networks - is Keras [4], an high-level framework capable of running on top of Google TensorFlow [5] which is an open source software library for numerical computation and Machine Learning applications.

3.1 Data preparation

The first important step prior to model creation is the data preparation. Every input data is given in a ROOT [6] TTree data format. This particular format is so widely used by the high energy physics community to be considered as a "standard" in relation to physics data analysis. ROOT files are also known for their ability to store from data tuples (TTrees) to histograms and

#	Optimizer	Training RMSE	Validation RMSE
1	AdaGrad	0.170369	0.169735
2	RMSProp	0.109598	0.110485
3	Adam	0.105400	0.107358
4	AdaMax	0.108883	0.110199

Table 1: Scores for the Neural Network training and validation phases.

N-dimensional plots. In order to deal with such a complex structure the best option available was to create a series of python macros able to correctly extract such information from the TTree structure and store it in a plain CSV (Comma Separated Values) file i.e. a format in which tabular data (numbers and text) are stored as a sequence of records each of which consisting of one or more fields, separated by a comma.

3.2 Neural Network structure

The simplest type of model is the Sequential model, a linear stack of layers. The first hidden layer has 1000 neurons and receives the information directly from the input layer of 27 different features with the sigmoid function selected as activation function and a gaussian distributed random number as kernel initializer (for the initial randomized weight assignment). Then a dropout is inserted: it consists in a random selection of a fraction of input units to 0 at each update during training time, preventing overfitting. The second layer is identical to the first one but contains only 50 neurons. In the end, the output layer (with only one node) closes the Network.

4. Results: comparison with Level-1 Trigger

Several models have been trained and tested and the best one was chosen for the comparison between the p_T predicted by Machine Learning and the p_T estimated by the p_T assignment unit of the Barrel Muon Track Finder. Table 1 shows the results - for both training and validation phases - used to choose the right model for analysis [7], in terms of Root Mean Squared Error (RMSE) loss, which is discussed as figure of merit elsewhere, e.g. in [8].

Figure 2 shows the p_T resolution for the entire range analysed (from 3 to 200 GeV/c), defined as the discrepancy of the predictions with the real generated value of p_T . The red line indicates the resolution of the current Level-1 trigger system while the blue line indicates the resolution of the predictions made by the Neural Network model. In particular, it is possible to notice a less broad distribution for the Machine Learning resolution, resulting in an overall improvement (yet small) with respect to the Level-1 trigger system.

Figure 3 shows the same p_T resolution histograms but in a specific momentum range (from 3 to 20 GeV/c). In general, a feature which is visible in all the p_T resolution plots is the scale: the peak of the ML-based distributions is, in fact, closer to zero than the Level-1 trigger peak. This is due to a different calibration of the p_T scale in the two cases, not considered during analysis. The Level-1 Trigger p_T assignment is in fact defined as the point for which the efficiency turn on curve reaches the 90% value. The ML, instead, attempts to estimate the generated muon p_T and is therefore characterized by a smaller scale bias.

Figure 4 shows an efficiency (or turn-on) curve: a threshold of 22 GeV/c was selected, as it is a

representative cut used during the data taking phase in Run 2. Such a momentum cut, in fact, has nearly full acceptance for signatures of events at the electroweak scale ($Z \rightarrow \mu\mu$, $W \rightarrow \mu\nu$) and above.

Figure 5 shows the same curve with a different p_T threshold (32 GeV). From these plots, a very high efficiency in the plateau region of the Machine Learning turn-on plot can be observed: the lower efficiency of the Level-1 curve in the high-pt region is mainly caused by the underestimation of the muon p_T in some events, which is visible as the small peak at -1 in the p_T resolution plots. The Machine Learning model, instead, is not affected by this issue and the efficiency at high values of p_T is nearly 100%. On the other hand, a small increase of efficiency in the 10 - 20 GeV/c region is noticeable, which is comparable with the worse resolution in the low p_T range (Figure 3). This is not optimal since it is located under the threshold thus resulting in a higher fraction of muons with low momentum misidentified as high momentum, and consequently in an increase of overall rate (yet to be evaluated).

Some aspects of this work will be analysed in more detail with further studies: an optimization of such models, including *ensemble* methods, has not been yet performed as well as their possible conversion into a high level synthesis firmware, in view of a future hardware integration with Field Programmable Gate Arrays (FPGA).

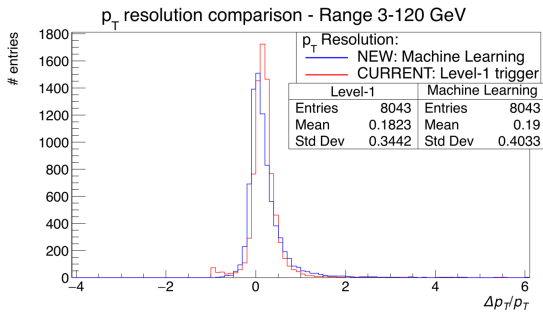


Figure 2: p_T resolution histograms computed for muons generated in the 3-200 GeV p_T range.

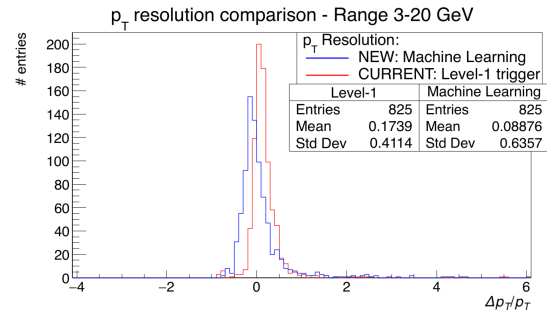


Figure 3: p_T resolution histograms computed for muons generated in the 3-20 GeV p_T range.

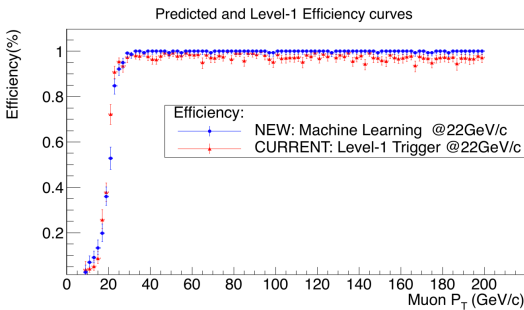


Figure 4: Efficiency turn-on given a p_T threshold cut of 22 GeV.

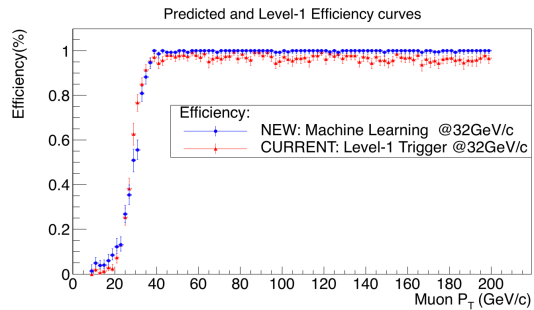


Figure 5: Efficiency turn-on given a p_T threshold cut of 32 GeV.

5. Conclusions

The precise evaluation of p_T is crucial to allow the trigger rate to be ultimately reduced and it represents a possible improvement of the CMS Level-1 muon trigger system during the High Luminosity phase of the LHC. After the evaluation of various linear and non-linear learning algorithms, an Artificial Neural Network approach was selected as the one offering the highest scores, in terms of the discrepancy between predictions and true values. Possible improvements - both in the software side, regarding the algorithm performances, and possible implementations on custom hardware - are being pursued in current developments.

References

- [1] CMS Collaboration, *CMS Technical Design Report for the Level-1 Trigger Upgrade*, Tech. Rep. CERN-LHCC-2013-011, CMS-TDR-12, CMS-TDR-012, CERN, Geneva, June 2013
- [2] J. Ero et al., *The CMS Level-1 Trigger Barrel Track Finder*, in 2016 *JINST* 11 C03038
- [3] S. Samarasinghe, *Neural Networks for Applied Sciences and Engineering: From Fundamentals to Complex Pattern Recognition*, CRC Press, 2016
- [4] Keras Documentation, URL: <https://keras.io>
- [5] TensorFlow Documentation, URL: <https://www.tensorflow.org>
- [6] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, in 1997 *Nucl. Instrum. Meth.* A389
- [7] T. Diotallevi, *CMS Level-1 Trigger Muon Momentum assignment with Machine Learning*, Master Degree Thesis, University of Bologna, July 2018 (Unpublished).
- [8] Ian Goodfellow et al., *Deep Learning*, MIT Press 2016.