Heavy flavour identification at CMS

Praveen Chandra Tiwari[*]

Research Scholar, Indian Institute of Science, Bangalore

E-mail: praveen.chandra.tiwari@cern.ch

Most of the CMS studies rely on the identification of b jets (b tagging), which is important for a broad range of analyses at CMS. Identification algorithms of jets from B hadrons heavily rely on machine learning tools and are thus natural candidates for advanced tools like deep neural networks. During the past couple of years, the CMS Collaboration has proven the power of deep neural networks implementing new algorithms, which outperform previous algorithms for b jet identification. While improving b tagging, the CMS Collaboration is pushing the heavy flavor identification beyond the traditional boundaries, with the implementation of b tagging algorithms specialized to the boosted topologies, and the development of c tagging algorithms, used to identify jets originated from charm quarks. With the increased experimentally excluded mass ranges of new particles, in several cases at the TeV scale, searches need to focus more and more on very boosted regimes. Several heavy flavor identification tools specific for boosted topologies have been developed to make these searches possible, such as b tagging of subjets and a double b tagger, aiming at the identification of boosted decays of the heavy particles into pairs of b quarks.

---

[*]On behalf of CMS Collaboration

PoS(ICHEP2018)898

1. Introduction

The CMS experiment at the CERN LHC requires the particles created in the LHC collisions to be reconstructed and identified as accurately as possible. With the exception of the top quark, quarks and gluons produced in pp collisions develop a parton shower and eventually hadronize giving rise to jets of collimated particles observed in the CMS detector[1]. Heavy-flavour jet identification techniques exploit the properties of the hadrons in the jet to discriminate between jets originating from b or c quarks and those originating from light-flavour quarks or gluons. A comprehensive summary of the newly developed and optimized techniques has been presented here.

2. Identification of heavy flavour jets

Algorithms for heavy-flavour jet identification use variables connected to the properties of heavy-flavour hadrons present in jets resulting from the radiation and hadronization of b or c quarks. The CSVv2 algorithm combines the information of displaced tracks with the information on secondary vertices associated with the jet using a MVA(multivariate analysis) technique and requires at least two tracks per jet. The DeepCSV algorithm, has been developed using a deep neural network with more hidden layers, more nodes per layer, and a simultaneous training in all vertex categories and for all jet flavours. The same tracks, IVF secondary vertices and other input variables are used in this approach as for the CSVv2 tagger, with only one difference, namely that for the track based variables up to six tracks are used in the training of the DeepCSV. The c jet identification algorithm uses properties related to displaced tracks, secondary vertices, and soft leptons inside the jets. Based on the presence or absence of a secondary vertex associated with a jet, three secondary vertex categories are defined in the same way as for the CSVv2 algorithm.

At the high centre-of-mass energy of the LHC, particles decaying to b quarks can be produced with a large Lorentz boost. When the decay of the boosted particle contains a b quark, b tagging can be applied either on the AK8 jet or on its subjets. In both cases the CSVv2 algorithm is used. In the first approach the CSVv2 algorithm is applied to the AK8 jet but using looser requirements for the track-to-jet and vertex-to-jet association criteria, consistent with the R = 0.8 parameter. In the second approach the CSVv2 algorithm is applied to the subjets. The limitation of the above approaches in identifying H$\rightarrow$ b$\bar{\text{b}}$ jets lead to a new double-b tagging algorithm, developed to fully use not only the presence of two b hadrons inside the AK8 jet but also the correlation between the directions of the momenta of the two b hadrons. Although the algorithm is developed using simulated H$\rightarrow$ b$\bar{\text{b}}$ events, any dependence of the algorithm performance on the mass or transverse momentum of the b$\bar{\text{b}}$ pair is avoided.

References

[1] CMS Collaboration, JINST **3**, S08004 (2008). doi:10.1088/1748-0221/3/08/S08004

[2] CMS Collaboration, JINST **13**, P05011 (2018) doi:10.1088/1748-0221/13/05/P05011

[3] CMS Collaboration, CERN-CMS-DP-2018-033, CDS: https://cds.cern.ch/record/2627468