B-Tagging through G.N.N

Cauê Evangelista

Sprace

Primary Analysis - Original Data

```
17-Dec-2022 01:19:32 CET  Initiating request to open file root://eoscms.cern.ch//eos/cms/store/relval/CMSSW_12_5_0/RelValTTbar_14TeV/GEN-SIM-RECO/125X_mcRun4_realistic_v2_2026D88noPU-v1/25
%MSG-w XrdAdaptor:  file_open 17-Dec-2022 01:19:35 CET pre-events
Data is served from cern.ch instead of original site eoscms
%MSG
17-Dec-2022 01:19:50 CET  Successfully opened file root://eoscms.cern.ch//eos/cms/store/relval/CMSSW_12_5_0/RelValTTbar_14TeV/GEN-SIM-RECO/125X_mcRun4_realistic_v2_2026D88noPU-v1/2580000/0
Begin processing the 1st record. Run 1, Event 46103, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:55.080 CET
Begin processing the 2nd record. Run 1, Event 46102, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:55.087 CET
Begin processing the 3rd record. Run 1, Event 46104, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:55.096 CET
Begin processing the 4th record. Run 1, Event 46101, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:55.099 CET
Begin processing the 5th record. Run 1, Event 46105, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:57.612 CET
Begin processing the 6th record. Run 1, Event 46106, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:57.616 CET
Begin processing the 7th record. Run 1, Event 46107, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:57.619 CET
Begin processing the 8th record. Run 1, Event 46108, LumiSection 462 on stream 0 at 17-Dec-2022 01:19:57.626 CET

[ParticleListDrawer] analysing particle collection pfTICL
```

idx	ID	Name	Stat	Mo1	Mo2	Da1	Da2	nMo	nDa	pt	eta	phi	px	py	pz	m	vx	vy	vz
0	22	gamma	0	-1	-1	-1	-1	0	0	0.366	-2.801	0.260	0.353	0.094	-2.999	0.000	0.000	0.000	0.000
1	22	gamma	0	-1	-1	-1	-1	0	0	0.306	-2.812	0.563	0.259	0.163	-2.537	-0.000	0.000	0.000	0.000
2	22	gamma	0	-1	-1	-1	-1	0	0	5.354	1.617	0.477	4.757	2.457	12.959	0.000	0.000	0.000	0.000
3	22	gamma	0	-1	-1	-1	-1	0	0	1.296	1.886	-2.245	-0.809	-1.012	4.175	0.000	0.000	0.000	0.000
4	22	gamma	0	-1	-1	-1	-1	0	0	0.366	2.548	-2.342	-0.255	-0.263	2.325	0.000	0.000	0.000	0.000
5	130	K_L0	0	-1	-1	-1	-1	0	0	0.948	-2.714	0.311	0.902	0.290	-7.119	0.140	0.000	0.000	0.000
6	130	K_L0	0	-1	-1	-1	-1	0	0	0.867	1.613	-0.470	0.773	-0.393	2.089	0.140	0.000	0.000	0.000
7	-11	e+	0	-1	-1	-1	-1	0	0	3.984	1.619	0.531	3.435	2.019	9.659	-0.000	0.000	0.000	0.000
8	11	e-	0	-1	-1	-1	-1	0	0	0.521	-2.755	-0.852	0.343	-0.393	-4.083	0.000	0.000	0.000	0.000
9	211	pi+	0	-1	-1	-1	-1	0	0	0.387	2.901	-1.336	0.090	-0.377	3.514	0.140	0.000	0.000	0.000
10	211	pi+	0	-1	-1	-1	-1	0	0	5.135	1.597	-2.160	-2.852	-4.270	12.153	0.140	0.000	0.000	0.000
11	-211	pi-	0	-1	-1	-1	-1	0	0	5.314	1.559	-2.211	-3.173	-4.263	12.075	0.140	0.000	0.000	0.000
12	-211	pi-	0	-1	-1	-1	-1	0	0	4.623	1.558	0.346	4.348	1.569	10.485	0.140	0.000	0.000	0.000
13	211	pi+	0	-1	-1	-1	-1	0	0	3.556	1.670	-2.448	-2.735	-2.272	9.109	0.140	0.000	0.000	0.000
14	211	pi+	0	-1	-1	-1	-1	0	0	2.787	1.923	-2.249	-1.749	-2.171	9.331	0.140	0.000	0.000	0.000
15	-13	mu+	0	-1	-1	-1	-1	0	0	2.372	1.583	-0.003	2.372	-0.007	5.530	0.106	-0.000	-0.018	-4.784
16	-211	pi-	0	-1	-1	-1	-1	0	0	2.215	2.658	1.775	-0.449	2.170	15.720	0.140	0.000	0.000	0.000
17	211	pi+	0	-1	-1	-1	-1	0	0	2.037	1.493	-2.206	-1.209	-1.640	4.305	0.140	0.000	0.000	0.000
18	-211	pi-	0	-1	-1	-1	-1	0	0	1.959	2.967	-1.286	0.550	-1.881	18.985	0.140	0.000	0.000	0.000
19	-211	pi-	0	-1	-1	-1	-1	0	0	1.801	2.042	0.870	1.162	1.376	6.825	0.140	0.000	0.000	0.000
20	211	pi+	0	-1	-1	-1	-1	0	0	1.786	-2.861	-2.875	-1.723	-0.471	-15.559	0.140	0.000	0.000	0.000
21	211	pi+	0	-1	-1	-1	-1	0	0	1.745	1.610	-2.271	-1.124	-1.335	4.190	0.140	0.000	0.000	0.000
22	211	pi+	0	-1	-1	-1	-1	0	0	1.677	2.383	-1.763	-0.321	-1.646	9.010	0.140	0.000	0.000	0.000

Primary Analysis - Filtering

Original Data

(41563,19)

Filtering

Separeted Events

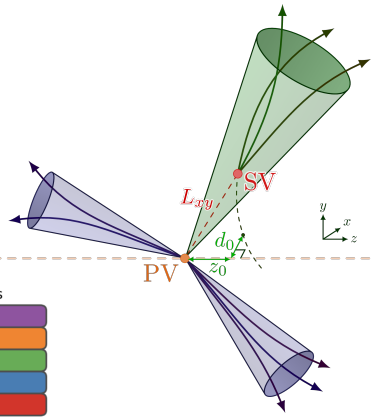
(1250,91,19)

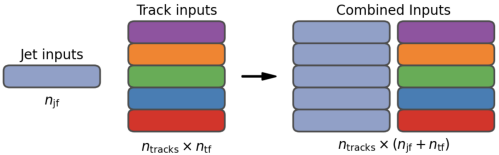
Next Steps - Events and Jet Clustering

- Events Array: Numpy $\rightarrow$ Awkward(eliminate zero padding)
  - Awkward: Awkward Array is a library for nested, variable-sized data, including arbitrary-length list.
- Jet Clustering: use FastJet from Scikit-HEP to perform Clustering
  - Scikit-HEP: is a project with the aim of providing the HEP community with an ecosystem for data analysis in Python.
  - FastJet: is a software package for jet finding in pp and e^+e^- collisions. This version includes bindings to Python and Awkward Array.

Next Steps - Tracks

- Just Particle Cloud it's not enough!
- How to simulate Tracks alongside with Jets?
- Which features are available to us? [1]





[1] Atlas Collaboration. *"Graph Neural Network Jet Flavour Tagging with the ATLAS Detector",2022.*

How? - Selection Criteria Features

Parameter	**Selection**
p_{T}	> 500 MeV
$\lvert d_0 \rvert$	< 3.5 mm
$\lvert z_0 \sin\theta \rvert$	< 5 mm
Silicon hits	≥ 8
Shared silicon hits	< 2
Silicon holes	< 3
Pixel holes	< 2

How? - Tracks Features

Jet Input	**Description**
p_{T}	Jet transverse momentum
η	Signed jet pseudorapidity
Track Input	**Description**
q/p	Track charge divided by momentum (measure of curvature)
$\mathrm{d}\eta$	Pseudorapidity of the track, relative to the jet η
$\mathrm{d}\phi$	Azimuthal angle of the track, relative to the jet ϕ
d_0	Closest distance from the track to the PV in the longitudinal plane
$z_0 \sin\theta$	Closest distance from the track to the PV in the transverse plane
$\sigma(q/p)$	Uncertainty on q/p
$\sigma(\theta)$	Uncertainty on track polar angle θ
$\sigma(\phi)$	Uncertainty on track azimuthal angle ϕ
$s(d_0)$	Lifetime signed transverse IP significance
$s(z_0)$	Lifetime signed longitudinal IP significance
nPixHits	Number of pixel hits
nSCTHits	Number of SCT hits
nIBLHits	Number of IBL hits
nBLHits	Number of B-layer hits
nIBLShared	Number of shared IBL hits
nIBLSplit	Number of split IBL hits
nPixShared	Number of shared pixel hits
nPixSplit	Number of split pixel hits
nSCTShared	Number of shared SCT hits
nPixHoles	Number of pixel holes
nSCTHoles	Number of SCT holes
leptonID	Indicates if track was used in the reconstruction of an electron or muon (only for GN1 Lep)

How? - Architecture