山东大学 SHANDONG UNIVERSITY

Department of Physics, Shandong University

Compressed EWK study(ISRC1N2)

Chengxin Liao
liaocx@ihep.ac.cn

Wed, Fri 19, 2024

Outline

1. Hyperparameters optimization
2. Performance of Model

Task-list

- Machine learning for LH channel
  - ~~check more Variable and select significance var for ML(DONE)~~
  - ~~BDTG hyperparameters optimization/ Setup a Grid Search framework (DONE)~~

- Preliminary study on multibody quantum mechanics (In Progress)
  QFT Lecture (Peskin part I)

- BSc thesis: https://www.overleaf.com/project/674e7119837a2580151a0868

- CS61A (python): https://cs61a.vercel.app/index.html

Hyperparameters optimization(HH)

Input(HH-Channel):

Sample:

Sig: ISRC1N2(mass_C1 = 100GeV, mass_N2 = 70GeV)->12180 entries
Bkg: 513850 entries

All input data(C1N2_100_70 and Bkg) already passed pre-selection

Strategy:

method: BDTG
Separate sig(bkg) into five folders, one for test, the other three for train, and last one for validation set,then traverse all possibilities.

```
Number of training and testing events
-------------------------------------------------------
Signal     -- training events            : 7311
Signal     -- testing events             : 2436
Signal     -- training and testing events: 9747
Background -- training events            : 308329
Background -- testing events             : 102770
Background -- training and testing events: 411099
```

Pre-Selection

$$had\text{-}had\ \ channel: nTaus \geq 2, nLeps = 0$$
$$pass\ MET\ trigger;\ MET \geq 200$$
$$1 \leq nBaseJet \leq 8$$
$$b - Veto$$
$$OS$$

Hyperparameters optimization(HH)

Variables:

Obj kinematics

Pt_tt

Angular correlations

dPhit1x
dEtatt
dPhiMax_xt
dPhiztt
dPhitt
dPhizxe
dPhiMin_xt
dPhit2x
dPhiMin_tj1
dRt2x
dRMax_xt
dRMin_tj
dRtt
sum_cos_dphi

Event kinematics

Mll(Invariant Mass of tau1 and tau2)
MIA
MT2_150
MET_Tau
Proj_tt
MstauA
MCT
frac_MET_tt
frac_MET_tau1
frac_MET_MeffInc_40
frac_MET_Meff

These vars are selected based on the importance

Hyperparameters optimization(HH)

Grid Search:

Ntrees: 200, 300, 400, 500
Max Depth: 6, 8, 10, 12
MinNodeSize: 1%, 2%, 3%
Learning Rate: 0.01, 0.05, 0.1

Binned significance: $Z = \sqrt{2((s_i + b_i)\log\left(1 + \frac{s_i}{b_i}\right) - s_i)}$

Show top Zn

	Model Name	Binned Significance	Max Zn	Max Zn Bin
500_12_1_005		14.2770	3.83857	199
300_10_1_01		13.9648	3.76965	198
200_6_1_01		13.9250	3.74940	198
500_6_3_01		14.2740	3.72616	199
400_10_1_01		13.9553	3.70167	199
300_6_2_01		13.9366	3.69620	199
300_10_2_01		14.0094	3.67743	199
300_8_1_01		14.0434	3.67624	198
200_8_1_01		14.1925	3.67005	198
400_12_1_005		14.1384	3.66529	199
200_6_2_01		14.2209	3.65978	199
200_6_3_01		13.7197	3.64427	199
500_10_1_01		13.8227	3.63722	198
500_8_1_01		13.8369	3.61405	198
400_10_2_01		14.2001	3.60950	199
500_6_1_005		14.0399	3.60132	197

Rebin result

```
Model Name,Binned Significance,Max Zn,Max Zn Bin,bin num
500_12_1_005,16.0862,3.8635,198,200
500_12_1_005,15.9967,3.62563,99,100
500_12_1_005,15.9318,3.62563,50,50
500_12_1_005,15.6612,3.07372,40,40
500_12_1_005,15.3086,2.45396,25,25
500_12_1_005,15.0825,2.20391,20,20
```

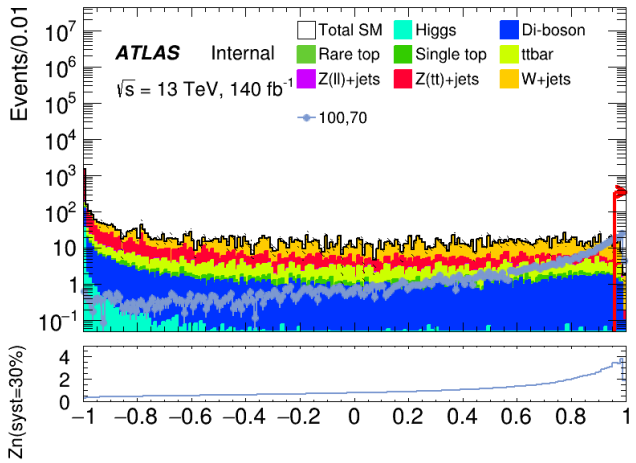
Compared with form result, there has a significant improvement in Zn

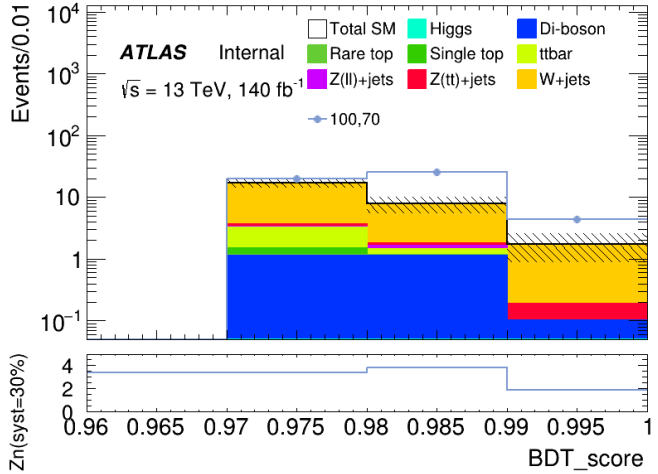
	Model Name	Binned Significance	Max Zn	Max Zn Bin
137	100_8_3_005	12.1380	3.27179	48
128	100_6_1_01	12.9663	3.24539	49
15	100_12_3_005	12.1254	3.22901	48
85	100_10_3_005	12.1150	3.20722	48
105	200_10_1_01	13.1608	3.19603	50
104	100_8_1_01	12.8853	3.18398	50
63	200_6_1_005	12.7673	3.17520	49
2	200_6_1_01	12.9052	3.17138	50
38	100_6_2_01	12.8248	3.16297	49
131	300_8_2_01	13.1256	3.16255	50
93	300_6_1_005	12.9703	3.14200	50
73	100_6_1_005	12.4457	3.14142	48
69	400_6_1_01	12.9285	3.14074	50
54	200_8_3_01	12.8685	3.13397	50
12	200_6_2_005	12.7035	3.12582	49
33	100_6_2_005	12.2453	3.11746	48
66	400_6_1_005	12.9369	3.10400	50
45	100_10_3_01	12.7388	3.10074	49
7	100_12_3_01	12.6318	3.10071	49
48	400_12_2_01	12.9393	3.09236	50
72	400_12_2_005	12.9179	3.06882	50
62	300_6_1_01	12.8501	3.06869	50
133	400_10_1_01	12.9846	3.06413	50
91	100_6_3_01	12.6291	3.06320	49
9	400_8_3_01	12.9337	3.06226	50
58	300_10_3_005	12.8854	3.05992	49
36	300_10_1_01	12.9906	3.05755	50
89	100_10_2_01	12.8406	3.05210	49
43	200_8_2_01	13.2400	3.04952	50
19	400_6_2_005	12.8263	3.04206	50

Performance of Model(HH)

Apply cut at BDT = 0.97

hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)





bin	max Zn	C1N2ISR (100,70)	bkg	Higgs	OtherTop	SingleTop	TopPair	VV	Wjets	Zlljets	Zttjets
198(0.97)	3.4089	**19.929+-0.858**	**17.186+-3.007(17.4%)**	0.007+-0.007	0.025+-0.014	0.345+-0.184	1.748+-0.528	1.125+-0.132	**13.445+-2.950**	0.180+-0.055	0.311+-0.085
199(0.98)	3.83857	**25.232+-1.287**	**7.805+-3.807(54.3%)**	0.000+-0.000	0.009+-0.017	0.003+-0.184	0.302+-0.570	1.138+-0.187	**6.034+-3.753**	0.116+-0.084	0.203+-0.114
200(0.99)	1.86049	**4.390+-1.350**		0.000+-0.000	0.000+-0.017	0.000+-0.184	0.000+-0.570	0.100+-0.189	**1.562+-3.850**	0.132+-0.139	0.086+-0.129

Hyperparameters optimization(LH)

Input(LH-Channel):

Sample:

Sig: ISRC1N2(mass_C1 = 100GeV, mass_N2 = 70GeV)->21225 entries
Bkg: 1703476 entries

All input data(C1N2_100_70 and Bkg) already passed pre-selection

```
Signal     -- training events            : 12735
Signal     -- testing events             : 4245
Signal     -- training and testing events: 16980
Background -- training events            : 1022092
Background -- testing events             : 340692
Background -- training and testing events: 1362784
```

Pre-Selection

$$lep\text{-}had\ \ channel:\ nTaus \geq 1, nLeps \geq 1$$
$$pass\ MET\ trigger;\ MET \geq 200$$
$$1 \leq nBaseJet \leq 8$$
$$b - Veto$$
$$OS$$

Strategy:

method: BDTG
Separate sig(bkg) into five folders, one for test, the other three for train, and last one for validation set,then traverse all possibilities.

Hyperparameters optimization(LH)

Variables:

Obj kinematics
- nBase_Jet
- mt_lep
- e_lep(energy of tau2)

Angular correlations
- dPhitt
- dRtt
- dRt1x
- dPhiMin_xj
- dPhiMax_tj

Event kinematics
- Mll(Invariant Mass of tau1 and tau2)
- METsig
- MT2_50
- Mwh(Invariant Mass of tau1 and MET)
- Mwl(Invariant Mass of tau2 and MET)
- MCT(Transverse Mass Squared)
- Proj_j(Projection of pt jet on zeta)
- Proj_tt(Projection of tau1+tau2 on zeta)
- mtx_tau
- Mtx_lep
- ht_tau
- mt_quad_sum
- mt_sum
- frac_MET_tau1
- frac_MET_tau2
- frac_MET_tt
- frac_MET_sqrtHT_40
- frac_jet_tau1
- frac_jet_tau2
- frac_jet_tt
- MT_taumin
- pt_Vframe

High importance at shiyi's feature

Note:

zeta is bisector direction of tau1 and tau2[PhyUtils::bisector(tau1, tau2)]

Hyperparameters optimization(LH)

Grid Search:

Ntrees: 200, 300, 400, 500
Max Depth: 6, 8, 10, 12
MinNodeSize: 1%, 2%, 3%
Learning Rate: 0.01, 0.05, 0.1

Binned significance: $Z = \sqrt{2((s_i + b_i)\log\left(1 + \frac{s_i}{b_i}\right) - s_i)}$

Show top Zn

	Model Name	Binned Significance	Max Zn	Max Zn Bin
50	500_12_1_005	16.0862	3.86350	198
105	200_10_1_01	15.9554	3.86006	198
138	400_10_1_01	15.8606	3.81102	199
93	400_12_1_001	15.4420	3.80592	192
119	500_12_1_01	15.7926	3.76671	199
42	400_10_1_005	15.9734	3.75629	198
133	500_10_1_005	15.9676	3.74424	198
142	500_10_1_01	15.7853	3.73933	199
4	200_10_1_005	15.8636	3.73165	196
57	300_10_1_005	15.9380	3.73121	197
140	400_12_1_005	16.0214	3.72224	197
0	300_12_1_005	15.9525	3.69489	197
72	200_8_1_005	15.7834	3.68745	196
80	400_10_1_001	15.3819	3.67258	192
24	500_8_1_01	15.6123	3.63555	199
13	500_12_1_001	15.5728	3.63147	193
121	400_12_1_01	15.8163	3.62834	199
26	300_10_1_01	15.7691	3.60172	199
83	500_8_1_005	15.8636	3.57064	198
79	300_12_1_001	15.1381	3.56449	188
3	400_8_1_005	15.8068	3.55725	197
117	300_8_1_005	15.7569	3.54771	197
137	300_12_1_01	15.7540	3.53785	199
82	200_12_1_01	15.8729	3.52960	198
12	400_8_1_001	15.2398	3.50279	191

Shiyi's result of LH channel

	hy	sig	zn
400_10_2_0.05		15.3225	4.47044
300_11_1_0.05		15.3127	4.65233
500_10_2_0.05		15.3099	4.32933
400_6_1_0.05		15.3075	4.69647
500_8_1_0.05		15.2990	4.30067
400_8_2_0.05		15.2980	4.49312
300_6_1_0.05		15.2929	4.91804
500_8_2_0.05		15.2891	4.35987
200_11_1_0.05		15.2849	4.71509
300_11_2_0.05		15.2804	4.62741
400_11_2_0.05		15.2780	4.42181
300_8_1_0.05		15.2753	4.59008
300_10_1_0.05		15.2733	4.46305
400_11_1_0.05		15.2701	4.33036
500_6_1_0.05		15.2593	4.61315
200_6_1_0.1		15.2559	4.69140
400_12_1_0.05		15.2554	4.29994
500_10_1_0.05		15.2493	4.19292
300_12_1_0.05		15.2461	4.45857
300_6_1_0.1		15.2424	4.48235
300_10_2_0.05		15.2319	4.52470
200_8_1_0.1		15.2283	4.44394
200_12_1_0.05		15.2279	4.68456
300_8_1_0.1		15.2267	4.28422
500_12_2_0.05		15.2216	4.18619
300_12_2_0.05		15.2161	4.43485
400_6_1_0.1		15.2142	4.28123
400_8_1_0.05		15.2139	4.29323

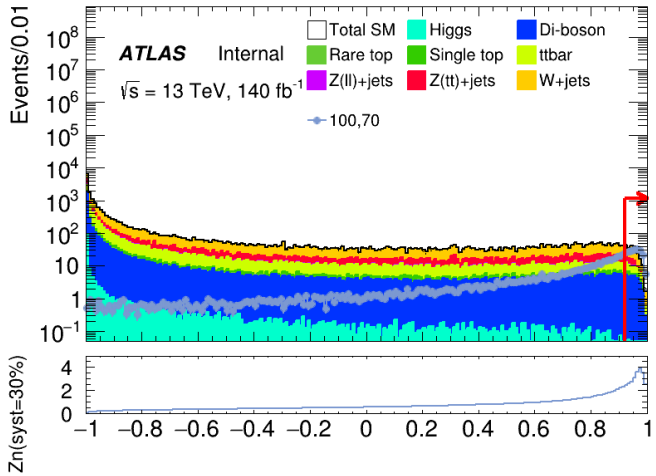
Rebin result

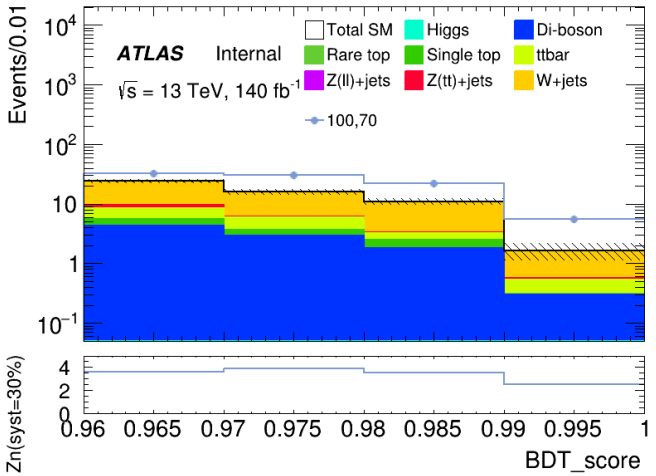
```
Model Name,Binned Significance,Max Zn,Max Zn Bin,Bin number
500_12_1_005,16.0862,3.8635,198,200
500_12_1_005,15.9967,3.62563,99,100
500_12_1_005,15.9318,3.62563,50,50
500_12_1_005,15.6612,3.07372,40,40
500_12_1_005,15.3086,2.45396,25,25
500_12_1_005,15.0825,2.20391,20,20
```

Performance of Model(LH)

Apply cut at BDT = 0.97

hyper parameter: NTrees=500, learning rate=0.05, max depth=12, MinNodeSize=1%(default)





bin	max Zn	C1N2ISR (100,70)	bkg	Higgs	OtherTop	SingleTop	TopPair	VV	Wjets	Zlljets	Zttjets
197 0.96-0.97	3.62563	32.151+-1.083	24.064+-2.244(9.32%)	0.038+-0.020	0.046+-0.023	1.325+-0.388	2.877+-0.663	4.210+-0.340	14.541+-2.075	-0.012+-0.044	1.049+-0.154
198 0.97-0.98	3.8635	30.244+-1.512	15.848+-2.885(18.2%)	0.006+-0.021	0.006+-0.023	0.801+-0.497	2.214+-0.875	2.911+-0.423	9.635+-2.663	0.051+-0.052	0.224+-0.196
199 0.98-0.99	3.53666	21.783+-1.754	11.019+-3.228(29.3%)	0.000+-0.021	-0.007+-0.026	0.731+-0.573	0.660+-0.937	1.811+-0.488	7.668+-2.988	0.012+-0.052	0.144+-0.222

TODO

1. add Mttau_min and pt_Vframe into ML

2. find a better way to select vars