Status of 4Top analysis

Huiling Hua[1], Fabio Iemmi[1], Hongbo Liao[1], Hideki Okawa[2], Yu Zhang[2]

1,IHEP, 2,Fudan University

April 7, 2021

Outline

1. Discussion on the strategy of MVA optimization
2. Discussion on the background estimation
3. Others

Strategy of MVA optimization

- Do data/mc comparison for all the input variables to avoid MC mismodeling.
- Provide the ranking of separation power and understand the ranking.
- Begin training with a minimum set of variables.
- Check the overtraining, correlation and the importance of the variables.
- Calculate the significance by $Z = \sqrt{2((N_s + N_b)log(1 + N_s/N_b) - N_s)}$. Make sure all the backgrounds are included.
- Remove variables with highly correlated with others but less separation power or add more variables.
- Require $N_S > 1$ after MVA selection? We can also design a loose category to increase statistic and sensitivity.
- Optimize the BDT parameters or try other ML methods. This is less important.

Background estimation

category	data	signal	$t\bar{t}$	SingleTop	TTX	SM Higgs	HH	total bkg	(data-bkg)/bkg
1Tau+0L	13942	9.474	6276.168	0.262	104.929	57.190	0.027	6438.577	1.165
1Tau+1L	2026	7.694	2329.306	0.651	65.438	35.281	0.003	2430.680	-0.166
1Tau+2L	78	1.843	62.112	0.555	19.325	6.488	0.002	88.482	-0.118
1Tau+3L	3	0.146	0.000	0.006	1.934	0.394	0.000	2.333	0.286
2Tau+0L	275	0.507	246.776	0.391	13.155	6.846	0.000	267.167	0.029
2Tau+1L	19	0.260	22.571	0.136	6.427	3.524	0.000	32.659	-0.418
2Tau+2L	1	0.031	0.271	0.004	0.729	0.354	0.000	1.358	-0.264

- It seems the non-prompt background is not worrisome.
- We will focus on the $t\bar{t}$ estimation in All categories and QCD estimation in 1Tau+0L.

MC-based estimation for $t\bar{t}$

- Top pt reweighting
  - Reweighting on Top pt at generator level.
  - Top PAG has a recommended correction.
  - It seems that ttH analysis derives this correction by themselves.
- Njet reweighting
  - The component of high jet multiplicity is underestimated.
  - A $t\bar{t}$+other-MC fit to data HT spectrum after pre-selection.
  - This is from full Run2 4top→1L AN.
- $t\bar{t} + bb$ reweighting
  - The component of $t\bar{t} + b\bar{b}$ is underestiamted.
  - Apply a scale factor of 1.2 on $t\bar{t} + b\bar{b}$ component of $t\bar{t}$ bkg. This number is from an updated measurement of $\sigma_{t\bar{t}+b\bar{b}}$.
  - Scale down other $t\bar{t}$ component to make total $t\bar{t}$ unchanged.
  - This is from full Run2 4top→1L AN.

Data-driven estimation

What we learn from 4top→full hadronic

- Estimate QCD+$t\bar{t}$ with ABCD method
  - Define CR by low N_{jet} and N_{bjet}
  - SR : $N_{jet} \geq 9$, $N_{bjet} \geq 3$
  - CR : $N_{jet} = 7, 8$, $N_{bjet} = 2$
  - 5 control regions, 1 signal region.
  - The normalization and the BDT distribution in SR can be derived.

We can define our own CR by reversing N_{jet}, N_{τ} or N_{bjet} selection.

Others(1)

- We need to fix the cross section ASAP.
- I am producing the Data/MC comparison plots for more variables. I will let you know when they are ready.
- Editing work on our AN.

Others(2)

- Our TopTagger is done before overlap removal of objects. This could be wrong. We need to be careful about the jet index saved in our ntuples.
- SingleTop Samples
  - We use $ST\_tW\_top, ST\_tW\_antitop, tZq\_ll, tZq\_nunu$
  - There are other samples ST_t-channel_top,ST_t-channel_antitop,ST_s-channel_4f_leptonDecays,ST_s-channel_hadronDecays