BDT Study And Combine Results Part1

Progress Report on Tau Final States of TTTT

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

[1]IHEP

[2]Fudan University

IHEP Group Meeting, 2021

Hypothesis test in combine

- $uS + B$: expected number of events
- Limit setting
    - $H_0 : u > 0$
    - $H_1 : u = 0$
    - expected limit: medium u limit under background only hypothesis
- Siginificance testing
    - $H_0 : u = 0$
    - $H_1 : u > 0$
    - expected signficance: under u=1 assumption, how much we can exlulde the background only hypothesis(medium significance)

From training to combine results

- apply the the training result to application, get the BDT distribution for all signal processes
  - use the same input variable list for application
  - run over the same samples for training(60% training, 40% testing) and applicaition(all events)
- feed signal and background distribution to combine
  - generate datacard(TMVApp_1tau1l_20var_datacard.txt)
  - text2workspace.py TMVApp_1tau1l_20var_datacard.txt -o TMVApp_1tau1l_11var_workspace.root
  - combine -M Significance TMVApp_1tau1l_20var_workspace.root -t -1 –expectSignal=1 –name TMVApp_1tau1l_20var (blind)
  - combine -M AsymptoticLimits TMVApp_1tau1l_20var_workspace.root –run blind –name TMVApp_1tau1l_20var (blind)

Datacard example

```
imax *
jmax *
kmax *
------------------------------------------------------------------------------------------------------------------------
shapes * *   /publicfs/cms/user/huahuil/TauOfTTTT/2016v1/TMVAOutput/v46_v2Resubmitv1/1tau1l_v2/AppResults/TMVApp_1tau1l_20var_forCombine.root  $PROCESS_MVA
_BDT
bin          SR_1tau1l
observation  -1
------------------------------------------------------------------------------------------------------------------------
bin        SR_1tau1l      SR_1tau1l        SR_1tau1l          SR_1tau1l            SR_1tau1l        SR_1tau1l         SR_1tau1l        SR_1tau1l
SR_1tau1l      SR_1tau1l      SR_1tau1l        SR_1tau1l          SR_1tau1l            SR_1tau1l        SR_1tau1l
SR_1tau1l        SR_1tau1l          SR_1tau1l            SR_1tau1l        SR_1tau1l         SR_1tau1l
process    WW             WGJets           ZGJetsToLLG        WWW                  WWZ              ZZZ               ttZJets          ttWJets      ttH
tZq_ll
           ST_tW_antitop  ST_tW_top        TGJets             THW                  THQ                                QCD_HT1500to2000
process    TTTT           TTTo2L2Nu        TTToHadronic       TTToSemiLeptonic     TTGJets
           0              1                2                  3                    4                5                 6                7
process    8              9                10                 11                   12               13                14               15           16
           17             18               19                 20                   21               22
rate       -1             -1               -1                 -1                   -1               -1                -1               -1           -1
           -1             -1               -1                 -1                   -1               -1                -1
------------------------------------------------------------------------------------------------------------------------
SR_1tau1l   autoMCStats   0
```
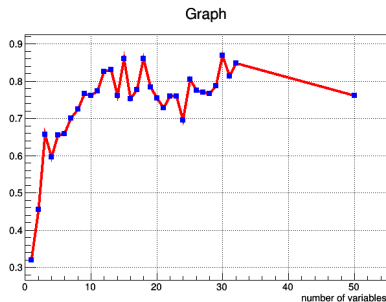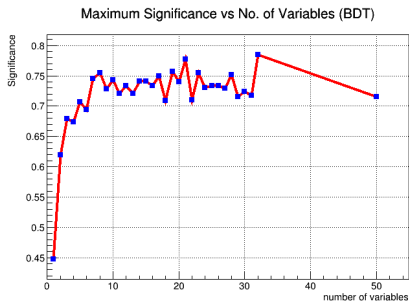
- no systemtic uncertainties considered yet
- have taken MC uncertainty into consideration
- removed empty processes from the card, due to the fact that combine doesn't accept empty histograms
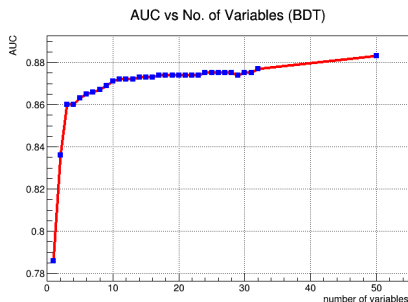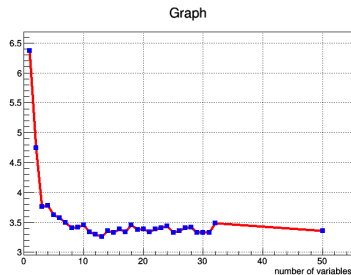
Combine results: expected significance

- left is maximum signficance, one bin counting experiment(calculated from test sample)(1tau1l)
- right is expected significance from combine, BDT shape based analysis(using application results)(1tau1l)

Combine results: expected limit

- left is expected limit, achieve plateau 3.5 at around 10
- right is AUC, achieve plateau around 10
- seems AUC is a pretty good estimator for BDT performance