Direct optimization of discovery significance

I. Pidhurskyi

Supervised by: A. Mohamed, D. Krüecker,I. Melzer-Pellmann, O. Turkot.

September 2019

Contents

1 Introduction

Supersymmetry (SUSY) concept was developed in the early seventies in the context of a string theory. At that time it appeared as a pure theoretical tool. Later on, it was realised that this symmetry may be a symmetry for four-dimensional space-time based quantum theories. Sins that time, lots of different models based on the supersymmetry were proposed. By providing boson-fermion mapping it naturally solves radiative corrections by canselations between fermion- and boson-loops, so that leading order (LO) terms remain dominant over the lower order terms; it also grants us with dark matter candidates, it could explain hierarchy problem, and many more [1].

Supersymmetry is known to be broken symmetry. But it means that we are not able to put constrains on masses for all the variety of particles it introduces. For this reason and the fact that no supersymmetry-particle has been observed yet (unless Higgs boson we detected is the lightest SUSY-higgs), we conclude that masses of these particles may be well beyond the energies available at modern accelerators. Also some of SUSY particles are expected to interact very weakly, like neutrinos, and so disabeling us to directly detect them. Nevertheless, these particles would still contribute to the deep inelastic scattering, and thus could be detected via indirect measurements.

This analysis aims at improvement of modern techniques for indirect studies of theories like SUSY, where signal is known to be highly dominated by various backgrounds and demands a state-of-art studies of kinematics.

2 Signal and background processes

This study is targetting the posibility to improve methods for SUSY analysis at LHC. Considered intercation is gluino pair production with each of them decaying to $t\bar{t}\chi_1^o$, where χ_1^o stands for the Lightest Supersymmetry Particle (LSP). The corresponding diagram is presented in Figure 1.

Main decay channel of the t ($\bar{t}$) quarks is the weak-decay channel to W^+ (W^-) boson and a b ($\bar{b}$) quark. Then there is two modes of W-decays: hadronic and leptonic. In this analysis we constrain to semi-leptonic events.

Resulting signatures for the signal events are: multiple jets originating from W-decays and 4 b-jets, a single lepton, and missing energy in transverse plane (MET) from ν and two χ_1^o. Number of jets originatin from hadronic W-decay depends on a boost of the W boson, so no tight selection can be performed for this quantity. Another problem is bad identification efficiency for b-jets, strong restriction on number of b-jets would dramatically reduce statistics.

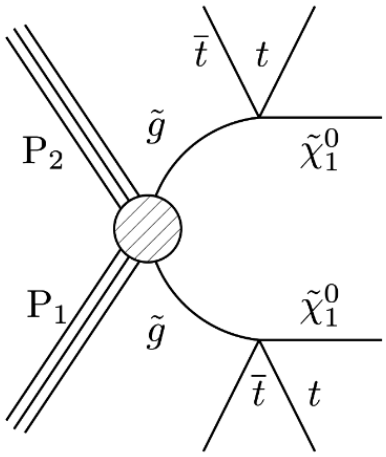

Figure 1: Diagram for signal events

Figure 2: tt-semileptonic

The main background for this case is *tt-semileptonic* (Figure 2). But due to possibility of misidentification of a lepton in W-decay, we will also obsereve a huge influence from *tt-dileptonic* (Figure 3) events.

Latest 95% C.L. limits are shown on Figure 4. Monte Carlo samples used for this study correspond to 35.9 fb^{-1} of data for 2016 year.

2.1 Event selection criteria

Following event selection was applied to the data. These criteria are set due to issues with trigger-system during 2016 run, which caused exclusion of these ranges from the data. Thus these cuts are set to match Monte Carlo samples to the available data.

- only events with a single lepton are selected
- lepton transverse momentum, $p_T^l > 25\text{GeV}$
- Number of jets with $p_T > 30GeV$, $N_{jets30} \geq 5$
- Second jet $p_T > 80$
- $H_T > 500$, where $H_T = \sum_{\text{Jets}} p_T^{\text{jet}}$

Figure 3: tt-dileptonic

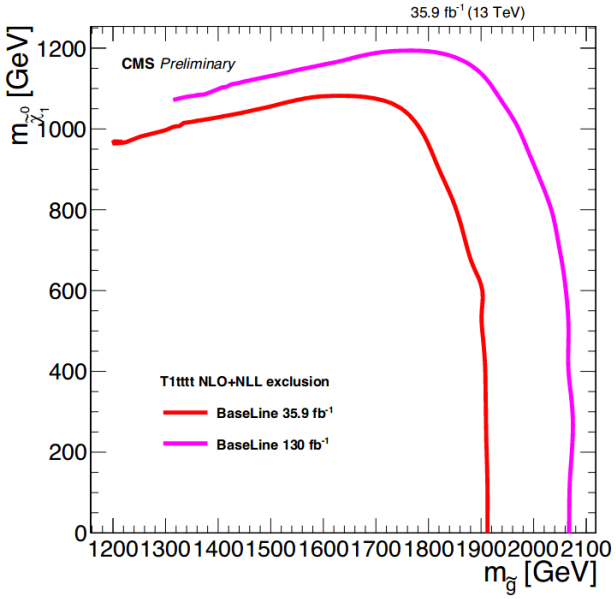

Figure 4: CMS Preliminary: 95% C.L. limits on $M_{\tilde{g}}$ and $M_{\chi_1^0}$

- $L_T > 250$, where $L_T = |p_T^l| + |\text{MET}|$
- $N_{b-\text{jets}} >= 1$

3 Deep neural network and loss function

Machine leaning (ML) techniques are getting used in high energy physics (HEP) analysis more and more today. They may dramatically simplify the studies by introducing a way to approximate signal without going too deep into details of physics behind the process, which sometimes becomes over complicated and demands enormous time to be studied well enough for a single analysis, which may even happen not to succeed. In context of HEP, it is perfect tool to handle analysis of event kinematics, and can be combined with the classic-approach analysis to obtain "cleaner" base to work around.

General idea of ML is to tune an over complicated functional to derive information of interest from a set of given features. For this reason, consideration on approach for such tuning might be crucial part of ML application. For most cases this tuning is implemented as minimization of some function, named *loss function*, which judjes the qualitative aspects of the output from the ML-model. In case of classification problems, common consideration for the loss function is cross entropy between the classes. Research of D.

Krüecker and A. Elwood ([2]) highlighted that this approach may not be an optimal choise for problem of signal identification in HEP. Instead of optimizing *accuracy* (approximated by cross entropy) of the classifier, one could actually consider to directly optimize discovery significance for the signal-class. Approximation of discovery significance is given by Asimov estimate (Equation 1).

$$Z_A = \sqrt{2\left((s+b)ln\left[\frac{(s+b)(b+\sigma_b^2)}{b^2+(s+b)\sigma_b^2}\right] - \frac{b^2}{\sigma_b^2}ln\left[1+\frac{\sigma_b^2 s}{b(b+\sigma_b^2)}\right]\right)} \quad (1)$$

And to turn maximization problem to minimization one, loss function is defined as $l_{Asimov} = 1/Z_A^2$, where square root operation is eliminated due to performance reasons.

Inheriting approach of previous study, we also consider deep neural network (DNN) approach.

4 Analysis setup

4.1 DNN structure

Following configuration for DNN-classifier was used:

- 2 fully connected hidden layers
- 256 nodes with ReLU activation in each layer
- single output with sigmoid activation

Output of the DNN is then treated as a probability of event to be a representative of signal or background classes.

4.2 Elimination of overtraining and consideration on optimizer

As deep neural networks are known to suffer from overtraining, 10% dropout [3] was applied for each layer. This method prevents DNN to "remeber" the exact definitions of signal and background representatives, and instead forces it to study the patterns of the classes. Yet, overtraining at some extent is still possible because training sample is likely to have some features specific to the particular sample (i.e. statistical fluctuations). To prevent DNN from such overtuning, validation-based early stopping [4] is applied. With this

approach, optimization will be stopped when no more improvement of DNN output for validation sample is observed.

Another important step is consideration on optimizer for training. In previous research [2], Adam [5] optimizer was used. It was observed that training with implemented loss function "explodes" at first iterations, unless model was pretrained with loss $l_{s/\sqrt{s+b}}$ [2]. In this study, such pretraining haven't solved the problem, and another optimizer had to be used instead. ADADELTA [6] algorithm has succeed with stabilization of the training with introduced loss, and thus it was considered for the study. Nevertheless pretraining was kept due to performance reasons, but loss for this stage was changed to cross entropy.

4.3 DNN input features

Folowing event parameters were chosen as input to the DNN:

- Missing Energy in Transverse plane (MET)
- $M_T = \sqrt{p_T^l p_T^{miss}(1 - \cos\varphi)}$, where p_T^l is a transverse momentum of the lepton, and p_T^{miss} is a missing transverse momentum
- p_T of a first and a second leading jets (two features)
- number of leptons in event (fixed by selection cuts)
- lepton transverse momentum (p_T^l)
- L_T
- H_T
- number of identified b-jets
- number of reconstructed $t/\bar{t}$-quarks
- N_{jets30}
- $\Delta\varphi$ – angle between lepton and reconstructed W boson momentum.
- relative isolation of the lepton
- lepton mini-isolation – ratio of the amount of measured energy in a cone to the transverse momentum of the lepton

5 Results

6 Conclusions

References

[1] Adel Bilal. Introduction to Supersymmetry. *arXiv e-prints*, pages hep–th/0101055, Jan 2001.

[2] Adam Elwood and Dirk Krücker. Direct optimisation of the discovery significance when training neural networks to search for new physics in particle colliders. *arXiv e-prints*, page arXiv:1806.00322, Jun 2018.

[3] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[4] Lutz Prechelt. Early stopping - but when? In *Neural Networks: Tricks of the Trade, volume 1524 of LNCS, chapter 2*, pages 55–69. Springer-Verlag, 1997.

[5] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv e-prints*, page arXiv:1412.6980, Dec 2014.

[6] Matthew D. Zeiler. ADADELTA: An Adaptive Learning Rate Method. *arXiv e-prints*, page arXiv:1212.5701, Dec 2012.