

Abstract

The ATLAS and CMS experiments at the LHC discovered a 125 GeV Higgs boson in 2012. The measurements of its properties give us an insight to many important physical parameters. In this thesis the CMS $H \rightarrow \tau\tau \rightarrow \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$ analysis is discussed, including the Higgs physics theory, construction of the CMS apparatus and event reconstruction algorithms. Moreover, a set of machine learning methods is presented together with their utilization in event identification. The construction of these methods allows them to be easily accommodated to official Monte Carlo CMS data samples in the supervised learning mode. We showed that the result of 0.875 for average area under receiver operating curves for all considered event final states is feasible. The best performance is obtained for neural network model.

The thesis contains the author analysis of significance implemented on top of both cut based and machine learning based distributions of discriminating variables in the aforementioned Higgs boson decay channel. The results indicate that the incorporation of the output of machine learning estimator can boost the performance of the analysis by 20 to 35%.

Streszczenie

W 2012 roku eksperymenty ATLAS i CMS ogłosiły odkrycie nowej cząstki skalarniej o masie około 125 GeV - bozonu Higgsa. Pomiary własności tej cząstki pozwalają na oszacowanie wielu ważnych parametrów fizycznych Modelu Standardowego. W szczególności sprzężenia bosonu Higgsa do leptonów mogą być mierzone w rozpadach na dwa taony. W niniejszej pracy omówiona jest analiza $H \rightarrow \tau\tau \rightarrow \mu\nu_\mu\nu_\tau\tau_h\nu_\tau$: przedstawiono podstawy teoretyczne Modelu Standardowego i spontanicznego łamania symetrii wraz z fizyką cząstki Higgsa, budowę detektora CMS przy akceleratorze LHC oraz algorytmy rekonstrukcji zderzeń. Zaprezentowano ponadto przegląd metod uczenia maszynowego razem z ich zastosowaniem do identyfikacji przypadków w omawianej analizie. Wygoda użycia tych metod wynika z faktu łatwego ich trenowania przy użyciu standardowych próbek Monte Carlo dostępnych dla analiz fizycznych kolaboracji CMS. Uzyskane wyniki wskazują na możliwość uzyskania wartości 0.875 dla uśrednionej powierzchni pod krzywymi ROC (*ang.* „Receiver Operating Curve”, charakterystyka operacyjna odbionika) dla rozpoznawania czternastu zdefiniowanych stanów końcowych istotnych w analizie $H \rightarrow \tau\tau$.

Praca kończy się autorską analizą znaczości zaimplementowaną dla wyżej wspomnianego kanału rozpadu bozonu Higgsa. Porównano tutaj rozkłady zmiennych dyskryminujących dla standardowej analizy opartej o cięcia kinematyczne oraz analizy opartej o wynik z modelu sieci neuronowej. Wykazano, że uczenie maszynowe pozwala na poprawę wyniku końcowego analizy o 20% do 35%.