DRF: Thesis SL-DRF-20-0650

RESEARCH FIELD

Theoretical Physics / Theoretical physics

TITLE

Machine Learning for Fundamental Physics

ABSTRACT

Context: Today, as a result of an outstanding experimental program, particle physics and cosmology are flooded with data. This offers us the unique opportunity to answer questions that have been the object of speculation for more than fifty years. In particular, particle colliders and Cosmic Microwave Background (CMB) surveys have collected a wealth of high-quality data in the past decades and may continue to do so for decades to come. Analyzing these datasets effectively is key to make progress in fundamental physics.

In particle physics, as more and more data are collected, the problems that confront us become sharper and harder to solve. We know that the theories that well describe current data are incomplete and should be extended, but our prior beliefs on how the extension should look like and on where to discover it experimentally become less concordant every day. This calls for a model-independent approach to data analysis.

In cosmology, well established measurements, such as the rate of expansion of the Universe from the CMB radiation, are being challenged by new observations. Reconciling these different observations or understanding if they signal the presence of new physical phenomena, can only be done by interrogating the data in novel ways. Mining these large, multivariate datasets for signs of new phenomena, presents many of the challenges that machine learning and deep learning are overcoming in a variety of other domains. In this project we leverage the unprecedented growth of these fields to shed light on the open questions of fundamental physics.

Thesis Project: We consider the problem of having large multivariate datasets that are seemingly well described by a reference model. Departures from the reference model can be statistically significant, but are

caused only by a very small fraction of events. The significance might stem from the extreme rarity of the discrepant events in the reference model and in this case anomaly detection techniques might be employed. Or the discrepancy is due to a small excess (or even a deficit) of events in a region of the space of physical observables that is also populated in the reference model. Our goal is to determine if the experimental dataset does follow the reference model exactly or if it instead contains “smal” departures as described above. In the latter case, we also want to know in which region of the space of observables the discrepancy is localized. This problem is relevant to Large Hadron Collider (LHC) datasets that are well described by the Standard Model of particle physics (SM) and CMB datasets that are well described by the standard cosmological model ?CDM.

I have already developed a new machine learning technique [1] that allows to analyze large datasets in a model-independent way, detecting data departures from a given reference model with no prior bias on the nature of the new physics responsible for the discrepancy. There are a number of potential applications in particle physics, astrophysics and cosmology. We are already taking the relevant steps in collaboration with CMS experimentalists to use this technique on LHC data. The next step is to apply this technique to CMB datasets in order to detect deviations from ?CDM (which can contribute to shed light on the current tension between different measurements of the Hubble parameter). The application of this technique to CMB data will be the main focus of the project.

In parallel we will refine [1] and look for the optimal model-independent new physics detection strategy.

This problem can be reduced to finding the minimum of a suitable loss function. Furthermore, the loss function first developed in [1] can be used to obtain better performances in traditional classification problems, with a variety of possible applications to fundamental physics (such as for example the separation between quark and gluon jets). There are many other directions that are worth exploring such as embedding [1] in a Generative Adversarial Network and using Autoencoders to compute a p-value rather than just signaling the presence of anomalous events. Depending on the progress of the student during his thesis project we will explore all these directions or only a subset, with the priorities outlined above: 1) CMB data mining 2) Optimal search strategy 3) Applications in classification 4) Variations on Autoencoders.

[1] R. T. D'Agnolo and A. Wulzer, Learning New Physics from a Machine, Phys. Rev. D 99, no. 1, 015014 (2019) doi:10.1103/PhysRevD.99.015014 [arXiv:1806.02350 [hep-ph]].

LOCATION

Institut de Physique Théorique
Service de Physique Théorique
Place: Saclay
Start date of the thesis: 01/10/2020

CONTACT PERSON

Raffaele D´Agnolo
CEA
Liste des pôles/Liste des départements/Liste des services/SPhT
IPhT, CEA/DRF
91191 Gif sur Yvette
Phone number: +33 1 69 08 66 30
Email: raffaele-tito.dagnolo@ipht.fr

UNIVERSITY / GRADUATE SCHOOL

Paris-Saclay
Physique en Île-de-France (EDPIF)

FIND OUT MORE

https://www.ipht.fr