

Sujet de thèse GTMP 2023

Titre : Développement de méthodes d'apprentissage statistique innovantes pour la recherche de signatures de nouvelle physique dans les données du détecteur ATLAS au LHC

Direction de thèse: Samuel Calvet et Julien Donini (LPC), Manon Michel (LMBP)

Contexte scientifique

Notre compréhension de l'infiniment petit s'appuie sur le Modèle Standard de la physique des particules. Bien que remarquablement performant, ce modèle est considéré comme une approximation d'une théorie plus complète qui reste à identifier. Grâce aux données exceptionnelles produites par le Large Hadron Collider (LHC) au CERN, de nombreuses théories alternatives ont été testées sans que leur validité ne soit démontrée.

Dès lors, il devient intéressant d'investiguer une autre piste, celle de la recherche d'anomalies. Il est en effet possible que les données contiennent des traces de nouvelles particules qui n'auraient pas été recherchées car aucune théorie testée jusqu'alors ne les avait prédites. La difficulté est alors de pouvoir explorer tout l'espace des possibles, ce qui pose le problème de la haute dimensionnalité, et ce sur des jeux de données de grande taille.

Projet de recherche

Le projet scientifique vise à développer et appliquer des méthodes d'analyse statistique modernes sur la quantité de données considérable collectées par le détecteur ATLAS pour la recherche de nouvelle physique au LHC. Deux directions principales seront explorées.

Tout d'abord, deux thèses réalisées au LPC en 2022 [1] ont montré le potentiel des réseaux de neurones non supervisés pour mettre en évidence de nouvelles particules dans un espace des paramètres donné. Maintenant que les preuves de concept de l'efficacité de ces méthodes [2] ont été établies, nous envisageons de les transférer aux données réelles collectées par ATLAS. Cela implique une mise à l'épreuve intensive de l'algorithme mis en œuvre, pour garantir que son application aux données ATLAS sera performante et ne souffrira d'aucun biais.

Ensuite, une autre approche consiste en une exploration globale d'un très grand nombre de signatures possibles de nouvelle physique. Cette stratégie, basée sur [3], est déjà en cours d'approfondissement par une collaboration entre l'équipe du LPC et des groupes de Tel Aviv et Montréal. L'enjeu ici est de pouvoir construire un jeu d'entraînement assez riche pour rendre possible la généralisation des performances obtenues aux données expérimentales. Une des techniques envisagées reposerait sur des méthodes d'apprentissage non supervisées couplées à une exploration de l'espace latent par marche aléatoire pour enrichir l'entraînement.

Ainsi, la thèse que nous proposons permettra de développer et d'exploiter ces méthodes afin d'analyser des types de collisions qui sont habituellement très difficiles à modéliser et donc pas ou peu regardés. En sciences les premières observations de nouvelles "régions" sont souvent riches en surprises, et dans le cas de la physique de particules la mise en évidence d'une nouvelle particule aurait des conséquences majeures.

Environnement scientifique

Les travaux envisagés pour cette thèse seront menés dans le cadre d'une collaboration entre le LPC et le LMBP.

L'équipe ATLAS du LPC est engagée depuis plusieurs années dans le développement d'outils d'apprentissage statistique innovants pour la recherche d'anomalies en physique des particules expérimentales. Avec trois thèses soutenues sur cette thématique, l'équipe du LPC s'inscrit dans une tendance toujours plus forte de l'utilisation de méthodes modernes d'apprentissage automatique pour la recherche en physique. Par ailleurs l'équipe a été chargée par le Centre d'IA pour la science du CNRS, AISSAI [4], de l'organisation d'une conférence internationale sur la recherche d'anomalie pour janvier 2024.

L'équipe de développement et d'analyse d'algorithmes stochastiques du LMBP apportera une expertise concernant les garanties théoriques, notamment en statistique Bayésienne et méthodes Monte Carlo, les améliorations possibles de l'efficacité des algorithmes employés et la mise en place d'estimateurs robustes de convergence. Cette équipe mène des projets sur des thématiques proches, comme les questions d'échantillonnage en inférence Bayésienne dans un contexte de physique des particules en collaboration avec l'expérience ATLAS (thèse A. Monemvassitis), l'analyse rigoureuse de la convergence de réseaux de neurones et le développement de nouvelles méthodes d'apprentissage non-supervisé avec comme visée des applications en cosmologie.

Références

[1] Thèses de Louis Vaslin (co-direction LPC/LIMOS) et Ioan Dinu (co-tutelle LPC-Bucharest)

[2] Documentées dans Rep. Prog. Phys. 84 124201

[3] <https://arxiv.org/abs/2107.11573>

[4] <https://www.cnrs.fr/fr/le-centre-artificial-intelligence-science-science-artificial-intelligence-aissai>