Snowmass Letter of Interest

Particle Physics and Machine Learning in Education

Mark Neubauer[1]

[1]University of Illinois at Urbana-Champaign, Champaign, USA

Abstract

The strong and growing role of machine learning (ML) in particle physics is well established and appropriate given the complex detectors and large data sets at the foundational layer of our science. Increasingly, Physics departments are offering curricula to their undergraduate and graduate students that focus on the intersection of data science, machine learning and physics. In this Letter of Interest, I provide some perspective on the potential role of particle physics in ML education and present some of the opportunities and challenges in the form of open questions for our community to explore.

1 Introduction

Particle physics holds a prominent role within academic curriculum. There are a number of reasons for this, including the "fundamental" nature of our science, the compelling historical develop of our field, theoretical research that applies and develops advanced mathematics, powerful applications, and high-visibility spin-off technologies (e.g. WWW).

At the same time, machine learning has an increasing prominent role in our science, as evident from any recent HEP conference and this Snowmass process. Physicists are increasing collaborating with computer scientists and industry to develop "physics-driven" or "physics-inspired" machine learning architectures and methods.

2 Opportunities

Physics departments are increasingly offering curricula to their undergraduate and graduate students at the intersection of physics, data science and machine learning. Particle physicists are increasingly interested in developing new courses at this intersection. For those so inclined, these courses provide opportunities for particle physicists to (1) describe synergies between modern machine learning research and particle physics research, (2) Make connections with colleagues from other departments, (3) make connections within your department in other research domains, (4) recruit students interested research at the intersection ML and particle physics, and (5) Learn ML ;)

3 Challenges

Among the primary challenges for particle physicists developing a machine learning course for their Physics Department are

1. **Not trying to do too much**. Our strengths lie in the analysis and interpretation of large scientific data sets and physics-inspired AI. Leave the foundational AI pedagogy to the CS courses

2. **Balancing physics and ML pedagogy**. Remember that its a physics course taught in the Physics Department. Its best to use as many physics examples and datasets to support your instruction as possible. Classification of jets and galaxies over cats and dogs.

4 My personal experience

A few years ago, I began discussing with our undergraduate physics majors at the University of Illinois at Urbana-Champaign about our curriculum in an attempt to assess the level of interest in a physics-oriented course in machine learning applications. The response was overwhelmingly positive and it was clear to me that many of our students want this type of training. In 2018, I developed a new course our Department titled *Data Analysis and Machine Learning Applications for Physicists*[1].

I designed this course to teach the fundamentals of scientific data analysis and interpretation to our students and empower them through practical utilization of modern machine learning tools and techniques using open source data and software. This course has a number of innovative technical elements and was, to my knowledge, the first course in the Physics Department to be delivered solely through Jupyter Notebooks, Git Hub, and mini research projects utilizing open scientific data.

I have taught this course to both undergraduate and graduate students over the Spring 2019 and Fall 2019 semesters, and will again teach PHYS 398MLA in the Spring 2021 semester. HEP postdoc Matthew Feickert and HEP grad student Dewen Zhong helped develop and deliver the course.

5 Outlook and Opportunities

In this letter of interest, we outlined some of the opportunities and challenges of developing curriculum at the intersection of physics and machine learning. A crucial element of these courses that involve student projects (such as [1]) is the availability and discoverability of *open scientific data for education*. Projects based on ML applications to physics data are a strength of these type of courses. As such, a well-curated set of data for education is critical. Is there a role for the Snowmass study data sets in education, training and outreach? It is also my hope that through the Snowmass 2021 community process we will curate a list of courses that HEP faculty have developed or taught. As machine learning for education, outreach and engagement is an integral part of Snowmass 2021, I would like to propose a small group effort to further discuss these topics.

References

1. M. Neubauer, "PHYS398 MLA". https://illinois-mla.github.io/syllabus/#course-description.