DEEP LEARNING FOR HIGGS BOSON IDENTIFICATION AND SEARCHES FOR NEW PHYSICS AT THE LARGE HADRON COLLIDER

EXECUTIVE SUMMARY

The Large Hadron Collider (LHC) is the world’s most powerful particle accelerator, designed to study the fundamental nature of matter and the forces that govern its interactions by colliding beams of protons at the highest-available energies. The research team is using Blue Waters to process, simulate, and analyze high-energy proton-proton collision data produced by the ATLAS experiment at the LHC and to improve researchers’ sensitivities to new phenomena by developing novel approaches to identifying Higgs bosons produced with high momentum at the LHC by using machine learning techniques.

RESEARCH CHALLENGE

The goal of particle physics is to understand the universe at its most fundamental level, including the constituents of matter, their interactions, and the nature of space and time itself. This quest is one of the most ambitious and enduring of human endeavors. The Standard Model (SM) of particle physics describes all known fundamental particles and their interactions, including the Higgs boson, which was discovered at the LHC in 2012 with significant contributions by the Illinois (Neubauer) Group. The discovery led to Francois Englert and Peter W. Higgs receiving the 2013 Nobel Prize in Physics. The SM has withstood the last 40 years of experimental scrutiny, with important exceptions being neutrino mass, dark matter, and dark energy. Recent developments in particle physics and cosmology raise the exciting prospect that we are on the threshold of a major step forward in our understanding.

It is an enormous challenge to process, analyze, and share the 15 petabytes of data generated by the LHC experiments each year with thousands of physicists around the world. To translate these enormous data sets through analysis and to search for new physics beyond the SM, the research team is using Blue Waters to implement this approach as a jet-mass identification tagger for ATLAS that can be applied to searches involving boosted dibosons and to search for new physics beyond the SM. Blue Waters resources are made available to the rest of the ATLAS collaboration in a variety of machine and deep learning approaches to improve measurements and to search for new physics beyond the SM.

RESULTS & IMPACT

Fig. 2a shows a particular one-month period in 2018 in which 35,000 Blue Waters cores were utilized to process 35 million collisions. The top panel of this figure shows that this approach is cost-effective, boosting cluster utilization, and has no adverse effect on other high-performance computing workloads. The job output was made available to the rest of the ATLAS collaboration for use in analysis of the LHC data to improve SM measurements and to search for new physics beyond the SM. Fig. 2b shows the Higgs boson identification accuracy as a function of the number of training epochs for a variety of DNN configurations and hyperparameter settings. The team is also using Hyperpar, a convenience wrapper using Hyperopt with Keras models, on Blue Waters to automate the scanning of hyperparameters in a variety of machine and deep learning approaches to improve the Higgs boson identification over backgrounds. The techniques show promise in addressing the challenges of boosted Higgs boson identification and improving the sensitivity of new physics searches at the LHC.

Figure 1: (Left) Decay of a heavy hypothetical particle X to a Higgs boson and a scalar particle that decays to a W boson pair. The decay products from this decay are highly collimated in the lab frame and therefore difficult to reconstruct using traditional methods. (Right) Minimum distance between final-state quarks versus Higgs boson transverse momentum. The dotted line shows the typical jet clustering radius, indicating that the majority of decay products are in close proximity to each other.

Figure 2: (Left) This shows a period of time during which 35 million ATLAS events were processed using 300 Blue Waters nodes. Utilization during this period averaged 81%, which is typical for Blue Waters. (Right) Higgs boson identification accuracy versus training epoch for the boosted Higgs boson tagger developed on Blue Waters.

WHY BLUE WATERS

Blue Waters, as a large CPU and GPU resource with high data-throughput capabilities, greatly facilitated this research. The strong support for containers allowed the research team to deploy their science application on Blue Waters’ nodes. Also, Blue Waters provided a means for a highly parallelized and automated scanning of free parameters in the team’s machine learning configurations and, therefore, rapid optimization of the researchers’ boosted Higgs boson identifier.

PUBLICATIONS & DATA SETS
