A Popular Science Summary of my Master's Thesis
Deep Autoencoders for Data Compression in High Energy Physics

Eric Wulff
Supervisors: Caterina Doglioni, Antonio Boveia, Lukas Heinrich

January 23, 2020

Artificial intelligence could help scientists in their search for new fundamental particles

A neural network trained to find patterns in experimental data from the world's largest particle accelerator, the LHC, might be able to help scientists in their search for new fundamental particles. In a preliminary study, it has been able to reduce the storage needed to save the enormous amounts of data produced by particle detectors such as ATLAS, located at CERN in Geneva, Switzerland.

Located at the European Organisation for Nuclear Research, CERN, in Geneva Switzerland is the world's largest particle accelerator, the Large Hadron Collider (LHC). Most of the time it is used to accelerate protons to speeds of up to 99.999999% the speed of light. That's just 10 km/h slower than light.

These protons are then smashed together in special collision points around the LHC. The proton-proton collisions result in a transformation of the collision energy into a myriad of new particles. Giant detectors are then used to record information like the mass, energy and paths of these new particles. This recorded information is what makes up the experimental data.

A special kind of neural network called an autoencoder was allowed to study some of the experimental data collected by ATLAS, the largest particle detector at CERN. The data used for this study consists of so called *jets*, which are basically groups of particles that all travel in roughly the same direction, and are a product of the highly energetic proton-proton collisions at the LHC. The task given to the neural network was to find patterns and correlations in the data that could be exploited to compress that data. If, for instance, the network would discover that two variables are related by a multiplicative factor it could save just one of the variables and remember this multiplicative factor. Then, when the second variable is needed, it can easily be computed from the first. In realty, of course, the relationships between jet variables can be much more complex, otherwise we wouldn't need to ask artificial intelligence to give us a hand.

Although this may sound great, there are still drawbacks. The network is never able to reconstruct the original data perfectly once it has been compressed. It can get very close but will never reach perfection. In some applications however, this doesn't have to be a problem. Certain analyses that scientists at CERN want to do do not require absolute precision but rather would benefit from having more experimental data. For such analyses, this neural network could be a big help for scientists and may contribute to discoveries of new physics in the future.