Unravelling the proton structure with artificial intelligence at the Large Hadron Collider

At the Large Hadron Collider (LHC) of CERN in Geneva, protons are accelerated to almost the speed of light and then made collide. By analysing the debris of these hugely energetic collisions, scientist study the laws of Nature at the smallest possible distance scales. The LHC represent the world's most powerful quantum microscope.

Protons are not elementary particles: they are composed by quarks, the proton building blocks, bound together by gluons. A precise knowledge of the structure of the proton in terms of its constituent quarks and gluons is necessary in order to interpret any measurement at the LHC. This makes possible studying the properties of particles such as the Higgs boson and searching for the new particles that make up the dark matter and have not been seen yet. This knowledge provides insights on the quantum theory of the strong nuclear force, Quantum Chromodynamics, for instance the size of antimatter inside the proton.

The NNPDF Collaboration (Amsterdam-Cambridge-Edinburgh-Milan-Singapore) has pioneered the use of machine learning methods to achieve this goal, and it currently produces the most widely used determination of the proton structure.

The collaboration has just submitted to the arXiv online repository two publications describing a new determination of proton structure, together with the release of the complete software analysis framework as open-source code (https://arxiv.org/abs/2109.02653 and https://arxiv.org/abs/2109.02671). Representing the collective efforts of the collaboration for over four years, this analysis is a major landmark in the ongoing quest to unravel the fundamental laws of matter at the smallest possible distances.

This new study of the proton structure benefits from a number of key improvements. State-of-the-art machine learning and artificial intelligence techniques are used to “learn” the underlying physical laws in a quasi-automated manner, with minimal human input. The constraints provided by a number of recent high-precision measurements from the LHC, HERA, and other experiments world-wide are exploited to impose stringent restrictions on the allowed models of the proton structure. Furthermore, theoretical calculations of improved accuracy make possible the robust interpretation of the available data.

The NNPDF4.0 determination is characterized by achieving unprecedented precision, with uncertainties down to the 1% level for a wide range of important LHC processes. Being able to satisfactory describe almost 5000 independent measurements within a single theoretical framework demonstrates the impressive success of our current understanding of the strong nuclear force. These results provide a crucial ingredient for many theory predictions for the upcoming Run III of the LHC, and then for its future high-luminosity run to start in 2027. In addition, the NNPDF4.0 analysis will make possible addressing key open questions in particle and nuclear physics, such as whether or not the proton contains heavy quarks, a hypothesis put forward four decades ago but still unresolved. Beyond particle physics, NNPDF4.0 is also a key input for applications in astroparticle physics, in particular to predict the outcome of the collisions of the highest energy particles of our universe, cosmic rays and neutrinos, with Earth matter and atmosphere.

Complementing these scientific results, the public release of the NNPDF machine-learning analysis code represents an important contribution to the global community of nuclear and particle physics. It will make possible a large number of spin-off applications covering a broad range of physics topics at the LHC as well as for the Electron Ion Collider, a new US-based accelerator which will study the proton, heavy nuclei, and the strong nuclear force with utmost detail.

With the release of its full analysis framework as open-source code the NNPDF collaboration is providing a significant contribution to the achievement of the open and FAIR (findability, accessibility, interoperability and reusability) science that is quickly becoming the paradigm within the global scientific community.