Contribution ID: **14** Type: **”Standard talk”**

RootInteractive tool for multidimensional statistical analysis, machine learning and analytical model validation

Wednesday 3 July 2024 15:00 (30 minutes)

RootInteractive is a general purpose tool for multidimensional statistical analyses, mainly used in the ALICE experiment at CERN. This Python-based tool enables dynamic, interactive visualisation and data aggregation and enhances capabilities on both the server and client side, expanding analysis possibilities for researchers and educators. As machine learning (ML) becomes increasingly important in multidimensional data analysis, interpreting ML models and assessing their uncertainties proves challenging, especially when the analyses reduce dimensionality, which can lead to misleading conclusions.

Our goal with RootInteractive is to streamline the management of complex multidimensional challenges. It allows users to visualise and fit multidimensional functions, incorporate uncertainty and bias, validate assumptions and define the functional composition of parametric and non-parametric analytical functions. This includes the use of symmetries and the development of multidimensional "invariant"functions/alarms, which are crucial for the validation of machine learning models and the optimisation of parameters in reconstruction, calibration and simulation.

RootInteractive uses a declarative programming paradigm that enables the construction of programme structures and the expression of computational logic without detailed control flow, making it easily accessible to professionals, students and educators alike. The tool supports interactive visualisation for both unbinned and binned data, facilitates n-dimensional histogramming/projection and enables the extraction of derived aggregated information on both the server (Python/C++) and client (JavaScript) side. It is compatible with client/server configurations via Jupyter and can also be used as a standalone client application or dashboard.

RootInteractive uses both lossy and lossless data compression and enables interactive analysis of large data sets —up to about O(10ˆ7) entries times O(25) attributes —in a compressed browser environment of about 500 MB. By applying representative downsampling (O(10ˆ-2 to 10ˆ-3)) followed by reweighting or pre-aggregation on servers or batch farms, it facilitates effective multidimensional analysis of monthly/annual ALICE statistics for calibration, reconstruction validation, QA/QC or statistical/physical analyses.

Recent development in RootInteractive have focussed on better integration of downsampling of representative data and support for conditional reweighting in interactive multidimensional aggregation. An important development was the integration of a domain-specific language that simplifies integration with RDataFrame and thus facilitates complex data operations.

Primary authors: IVANOV, Marian I (GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE)); IVANOV, Marian (Comenius University (SK))

Presenter: IVANOV, Marian I (GSI - Helmholtzzentrum fur Schwerionenforschung GmbH (DE))

Session Classification: Plenary Session Wednesday