LA-UR-18-30086
Approved for public release; distribution is unlimited.

Title: Finding Outliers in Differential and Integral experiments using Machine Learning Techniques

Author(s):
Grechanuk, Pavel Aleksandrovi
Grosskopf, Michael John
Haeck, Wim
Herman, Michal W.
Neudecker, Denise
Rising, Michael Evan
Vander Wiel, Scott Alan

Intended for: ND2019, 2019-05-19/2019-05-24 (Beijing, China)
Web

Issued: 2018-10-23

Finding Outliers in Differential and Integral experiments using Machine Learning Techniques

P. Grechanuk, M. Grosskopf, W. Haeck, M. Herman, D. Neudecker, M. Rising, S. Vander Wiel

One major task of nuclear data validation is finding outliers in calculated over experimental (C/E) values of integral benchmark experiment and pinpointing whether outlying C/E values are caused by shortcomings in experiments or in nuclear data. If an imperfect prediction of a benchmark experimental value is attributed to nuclear data, then the next question is: which nuclear data observable needs to be improved? Another problem is identifying outlying experimental data for a nuclear data evaluation and discarding the appropriate ones. This problem is particularly difficult if there are two or more groups of incompatible experimental data and one has to choose a group for a reliable evaluation. The core problem of those two examples is the same: One wants to understand the potentially complex relationship between outlying experimental values and features in measurements and/or nuclear data. Machine learning methods like clustering and random forests can be used to learn and understand these relationships.

Here, we apply these techniques to identifying the reasons and groups of outliers for two classes of data. The first class are the C/E values for selected ICSBEP critical assemblies using measurement and nuclear data features. The second class of experiments are differential ^{239}Pu(n,f) cross-section measurements used for the Neutron Data Standards evaluation. Using these two example, we cross-compare the feasibility of applying machine learning techniques to these typical nuclear data problems and show how reliable the classification is dependent on number of experiments and features.