Application of Machine Learning algorithms for experimental data processing and estimation of ^{96}Mo(n, p)^{96}Nb reaction cross section

Sangeetha Prasanna Ram[1], Jayalekshmi Nair[1], Singh Vivek Virendra[2], Dhanamma Jagli[2], S. Ganesan[3], S. V. Suryanarayana[3]

[1] Department of Instrumentation, V. E. S. Institute of Technology, Mumbai, India 400074

[2] Department of Master of Computer Application, V. E. S. Institute of Technology, Mumbai,India 400074

[3] Former Scientist, Nuclear Physics Division, Bhabha Atomic Research Centre, Mumbai, India 400085

Abstract

Nuclear data available in articles and research papers, goes through the nuclear data pipeline process comprising of acquiring published data, compiling, evaluating and processing them into high quality data prior to applying them in various fields such as nuclear medicine, nuclear power, simulations and many more. Among the various steps in the nuclear data pipeline, nuclear data evaluation is the process of combining the experimental and theoretical model data for determining the optimally recommended value of nuclear data such as reaction cross section, along with its uncertainties, over a large incident neutron energy range [1]. However, experimental data available in IAEA-EXFOR database library [2], entered by experimenters based on the nuclear experiment performed by them, do not cover the entire range and are sparsely distributed. This results in the nuclear data evaluators to largely depend on theoretical model data for nuclear data evaluation. To overcome this, the nuclear data evaluators work on the analysis of the experimental data and its covariance to estimate its values at the energy points where actual experimental data is unavailable.

In this study, for the first time, estimation of experimental data of ^{96}Mo(n, p)^{96}Nb reaction cross section data at neutron energies where actual data is not available, is performed using Machine Learning (ML) regression algorithms, as shown in Fig 1. This has been performed by initially studying all the EXFOR papers corresponding to ^{96}Mo(n, p)^{96}Nb reaction from EXFOR database in order to understand the facility and procedure employed during the nuclear data experiment. Then based on the detailed study inclusion of an experimental data in the nuclear evaluation was decided and outliers were identified. At the same time, a Python code for cleaning the data and for identifying the outliers was also developed and the outcome of both the methods were compared and found to be satisfactory. Also, some of the experimental data was renormalized by considering the latest values of the attributes.

Further to the cleaning and preparing of nuclear experimental data, it was then split as training and testing data in the ratio of 80:20 and an estimation model was developed based on training data set. Three popular simple linear regression ML algorithms namely Ordinary Least Square (OLS), Least Absolute Shrinkage and Selection Operator (Lasso) and Ridge regression [3], were employed in this study, for developing the model. Lasso and Ridge regression techniques, which are less influenced by large outliers unlike OLS technique, are improvisations of least square regression analysis. These regression techniques perform both variable selection and regularization for improving the prediction accuracy. The training data points of ^{96}Mo(n, p)^{96}Nb reaction cross section along with its uncertainties, were used by the ML algorithms, to learn and develop the model parameters. The optimal model was selected based on the minimum root mean square error for prediction. The selected model was then subjected to the test data and based on its prediction accuracy, tuning of the model hyperparameters was performed to get the best results [4].

Finally, the model was validated based on unseen data and the model giving the best prediction accuracy was selected as the most suitable model. For this study of developing the regression model for experimental data of ^{96}Mo(n, p)^{96}Nb reaction cross section, both Lasso and Ridge regression techniques were found to be giving marginally better results than the OLS regression technique.

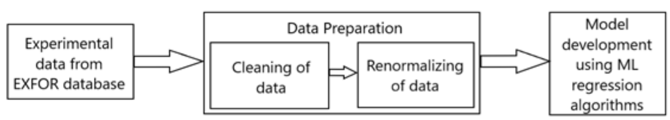

Fig. 1: Experimental data processing and estimation using Machine Learning algorithms

References

[1] Donald L. Smith (1996), *Some guidelines for the Evaluation of Nuclear data*, BNL-NCS-63006.

[2] IAEA-EXFOR Database available at http://www-nds.iaea.org/exfor.

[3] Frank Emmert-Streib, Matthias Dehmer (2019), *"High-Dimensional LASSO-Based Computational Regression Models: Regularization, Shrinkage, and Selection, Machine Learning and Knowledge Extraction"*, Vol. 1(1), 359-383.

[4] Sangeetha Prasanna Ram, Jayalekshmi Nair, et al., *"Application of Kalman filtering technique for Evaluation of neutron cross section data of 100Mo (n, 2n) 99Mo reaction"*, accepted for publication in - Nuclear Instruments and Methods in Physics Research Section A, Article No. 165850.