Approved for public release. Distribution is unlimited

ORNL/TM-2020/1535

WANDA: AI/ML for Nuclear Data

Vlad Sobes (UTK)
Michael Grosskopf (LANL)
Kyle Wendt (LLNL)
David Brown (BNL)
Michael S. Smith (ORNL)
Patrick Talou (LANL)

5/01/2020

OAK RIDGE NATIONAL LABORATORY
MANAGED BY UT-BATTELLE FOR THE US DEPARTMENT OF ENERGY

ORNL/TM-XXXX/XXX

ORNL Physics Division

WANDA: AI/ML for Nuclear Data

Vlad Sobes (UTK), Michael Grosskopf (LANL), Kyle Wendt (LLNL), David Brown (BNL), Michael S. Smith (ORNL), Patrick Talou (LANL)

Date Published:
5/1/2020

Prepared by
OAK RIDGE NATIONAL LABORATORY
Oak Ridge, TN 37831-6283
managed by
UT-BATTELLE, LLC
for the
US DEPARTMENT OF ENERGY
under contract DE-AC05-00OR22725

WANDA: AI/ML for Nuclear Data

Draft Summary of AI/ML Session
Workshop on Applied Nuclear Data Activities 2020
Washington, DC
March 3 – 5, 2020

Session Chairs: V. Sobes (UTK), M. Grosskopf (LANL), K. Wendt (LANL), D. Brown (BNL), M. S. Smith (ORNL), P. Talou (LANL)

Rapporteurs: A. Lewis, P. Vicente Valdez

Presentation Contributors: A. Lewis, M. Kuchera, L. Neufcourt, P. Vicente Valdez, J. Escher, S. Yoo, D. Neudecker, J. Hutchinson, N. Schunk, A. Lovell, G. Zhang

Introduction

The fast spread and impact of machine learning (ML) and artificial intelligence (AI) models to diverse areas of physical sciences indicate their tremendous potential to address critical issues and potential bottlenecks in the nuclear data pipeline. The nuclear data community has identified a number of key areas in which AI/ML advances already show substantial promise to make significant impacts, both in the short term and long into the future. By leveraging advances made in other areas of science and, simultaneously, driving innovations in AI/ML, we can address the needs of the community to provide more rapid, accurate, and robust evaluations, quicker compilation of both data and critical contextual information from published experimental work and optimal experimental design for validation. Additionally, AI/ML surrogate models may be used to incorporate more representational and realistic physics models into transport code simulations, as well as rapidly reproduce the results of complex multi-physics codes relevant for a wide range of applications. Targeted investments are needed now to fully realize the potential of AI/ML in nuclear data.

A critical portion of this investment should be directed to fostering collaborations between nuclear researchers and experts in the AI/ML community, especially within the Department of Energy labs and universities. Through collaborations, more appropriate algorithms for solving critical research problems may be most efficiently determined and subsequently trained, tuned, and deployed for maximum scientific impact. While open source tools like TensorFlow, Keras, PyTorch, and scikit-learn are valuable because they can be used for rapid exploration of new and innovative ideas, in the hands of non-experts, they can yield biased results with unphysical properties.

There are other areas for investment as well. To enhance the reproducibility of results, and to best leverage advancements across different areas of the pipeline, efforts to collect and share fitted models and data should be explored. By including notes on applicability and limitations of different AI/ML approaches, these collections may be more robust. Another focus area is on developing rigorous approaches for validating the trustworthiness of AI/ML methods in the nuclear data context, as these validations will be necessary before these tools can be deployed in certain applications where safety is paramount (e.g., nuclear energy, nuclear security). The integration of uncertainties arising from the use of AI/ML tools into the larger uncertainty quantification (UQ) process in nuclear data could be an essential ingredient of future validation efforts. Finally, advances in AI/ML should be considered a method for augmenting nuclear data expertise – to supplement the judgement and work of experts, not replace them.

Evaluations and Processing

Modern evaluations are built from a collection of disparate phenomenological models which are fit to experimental data. Ideally, the best possible theoretical models of the underlying structure, reaction and fission physics would be included. However, while capable of capturing and predicting trends across the nuclear chart (i.e., correlations between observables), these theoretical models are computationally intensive and often lack the descriptive power and accuracy needed to reproduce experimental data. The consequence is that evaluation models have limited predictive power, which leads to progressively lower-quality evaluations for unstable or difficult-to-measure nuclei where little or no data exists. Machine learning has the potential to address this issue on multiple fronts. Emulators have already proven their usefulness in providing tremendous speedup of fundamental DFT and QMD calculations for mesoscopic material science and chemistry. Such emulators could be used to address the extreme computational expense and permit the partial inclusion of models into modern evaluations. While present theoretical models could be insufficient for evaluations where a great deal of data exist, AI/ML may significantly enhance extrapolations to regions of the nuclear chart where little or no data exists. Further, theoretical models will necessarily impart more physically correct correlations between observables, both in the sense that correlated observables are better described, and that the extracted covariance matrices better describe the relationships between different processes.

The applications of machine learning continue from the specific nuclear reaction into validating the evaluation libraries in their entirety with respect to integral experiment and into processing those libraries for applications. Insight into the defects and missing important physics can be gained by studying libraries. Tools such as unsupervised learning have the potential to help identify new systematic trends in the nuclear data evaluations that may have been missed by human evaluators, and be a critical aid for enhancing or correcting our models and methodology. Even in cases where we cannot correct our models, it opens the possibility to better account for their defects and avoid overfitting. Machine learning may also enhance how we post-process and encode nuclear data for applications. Two areas ripe for study are compressing post-processed libraries into a memory-limited form and building better/more adaptive multigroup cross sections. This is particularly important as high-performance computing centers move toward more processing power with less memory.

Experiments / Compilations

The compilation and analysis of experimental data is a crucial step in the nuclear data evaluation pipeline. Without useful, accurate, well-documented and vetted experimental data sets, the resulting evaluated files can lead to significantly erroneous and biased results that are further propagated to applications – with possibly disastrous consequences. The EXFOR experimental data library constitutes a unique and valuable resource for the nuclear data community. However, in its current form, it does not yet satisfy all the needs of evaluators, especially concerning automatic reading of large amout of data and ML-supported interpretation thereof, a situation that is exacerbated if modern ML/AI algorithms are to be unleashed on it.

To make large-scale machine learning with reaction data possible, the community will need to develop a new database to store the data sets that have been vetted, standardized, and – in some cases – adjusted. The final version of the data sets in this database should be standardized (in formatting, metadata tags and uncertainties), quality-verified (checked for compilation errors and experimental biases and updated to new standard and structure values), and well-characterized (uncertainties standardized using experimental uncertainty templates). The compilation of data into this new database should be based on communication between the evaluator or qualified data users, the original compilers, and the authors of the published work to ensure that the highest quality database is created. To streamline data vetting, natural language processing (NLP) tools need to be developed to simplify the correction of compilation errors, and available data quality verification software should be utilized.

With targeted investments, NLP compilation of nuclear data may become the norm. However, there are many challenges to automated NLP, e.g., processing errors and issues with PDF files. While there have been recent developments with NLP tools for other scientific fields, further development is needed for table and figure processing, for incorporating a lexicon of nuclear data terms and symbols, and for developing an interface for expert validation to leverage automated NLP processing capabilities in nuclear data.

Validation

Validation of nuclear data is a key pillar in the nuclear data pipeline, and the potential for AI/ML to impact the validation process is significant. AI/ML has the potential to process complex relationships in the large spaces spanned both by nuclear data and simulated integral experiments. It may be invaluable in directing the design of optimal validation experiments to probe the potential deficiencies in ND models and evaluation, utilizing iterative feedback in the estimation process. Additionally, the incorporation of information about the inability to predict observations in benchmarks across applications and utilize this information holistically using AI/ML may reduce potentially compensating errors due to adjustment to a particular application. This approach will require the sharing of benchmark data and metadata across application domains in a way that allows relationships between nuclear data and discrepancies to be captured by AI/ML, while respecting potential limitations due to information sharing related to security and intellectual property.

To build toward that goal, advances can be made now in both AI/ML-guided optimal experimental design and ML-guided search for deficiencies in nuclear data estimates with respect to integral experiments. The areas of sequential design/optimization using Gaussian processes and deep neural networks show exciting promise to couple with physics models for searching a wide space of experimental designs. Additionally, advancements in reinforcement learning show tremendous potential in sequential decision-making tasks that may be applicable in this area as well. ML prediction models are well known to capture complex relationships between input features to the model and the target of prediction to obtain impressive accuracies. Utilizing tools for ML interpretability, the relationships learned by the model can be communicated to nuclear data experts to allow for directed investigation into potentially unexpected deficiencies in nuclear data.

Synergistically, the expertise developed over decades for nuclear data validation can also drive developments of approaches to validation of AI/ML methods that are desperately needed in this community and beyond. Given the importance of nuclear data in safety and security applications, building trust in the stability and robustness of results obtained through deployment of AI/ML methods is critical. Advancements in interpretability ensures that the way in which the AI/ML results are obtained avoid relying on spurious relationships in the data, while showing results are stable to variations in data, to the ML fitting process, and across implementation in the community will ensure that AI/ML results are trustworthy.

Applications

Specific applications, such as detector model responses or correlated signatures of nuclear physics processes (e.g., fission) cannot rely solely on average quantities as tabulated in ENDF/GNDS-formatted libraries. Instead, they require the incorporation of complex physics models into transport simulations, significantly enhancing the computational cost of such calculations. Emulators developed from these more fundamental calculations could be developed and render such simulations tractable, opening a new frontier for transport code capabilities. Such developments would complement, not replace, the existing tabulated nuclear data libraries.

Another application of interest is the identification of features and hidden patterns in the complex, highly multi-dimensional and correlated phase space of nuclear data, especially as used in the simulation of various nuclear applications, including critical and sub-critical assemblies, pulsed spheres, radio-chemistry, reactor designs, and others. ML algorithms associated with well-structured databases of nuclear data and nuclear applications provide an unprecedented opportunity to pull out features and parameters in a comprehensive manner, providing feedback on data evaluations from a wide range of diverse applications. This is an extremely difficult task for individual researchers who are often experts in only pieces of this giant puzzle, and who tend to find practical solutions to problems of limited scope, which can in some cases lead to compensating errors in the nuclear data that negatively impact other applications.

AI/ML methods in the development/training of surrogate models also have the potential to contribute to a range of nuclear data applications. First, fast surrogate models have the potential to significantly increase the computational capacity for fast propagation of uncertainties through multi-physics problems. In particular, multi-scale (multi-fidelity) approaches are promising, where several surrogate models, differing in fidelity and speed, can be combined to cover a wide range of simulations. Furthermore, such multi-scale surrogate models can be used to study a very wide design space in the optimization of integral experiments. By combining these surrogates with targeted full-fidelity simulations, the development of on-the-fly learning algorithms may be possible.

Conclusion

AI/ML approaches have tremendous potential to address critical short-term and long-term needs across the nuclear data pipeline. To realize this potential, we have indicated the urgent need for targeted investments to leverage advances made in other areas of science and, simultaneously, drive innovations in AI/ML. Critical areas for investment include: fostering collaborations between nuclear data researchers and experts in the AI/ML community; collecting and sharing fitted models and data along with the relevant notes on applicability and limitations; developing approaches to validate the trustworthiness of AI/ML methods; assessing UQ contributions of these methods; and developing surrogate models to rapidly emulate the results of complex multi-physics codes. Throughout these efforts, the principle should be followed to develop AI/ML tools that augment nuclear data expertise, not replace it, and to ensure that, when necessary, rigorous safety protocols are followed.