IAEA
International Atomic Energy Agency

Summary of the Working Group on AI for Nuclear Data

WG Coordinators

Christian Hill, Ludmila Marian, Georg Schnabel

Technical Meeting on Artificial Intelligence for Nuclear Technology and Applications
#AI4Atoms Virtual Event
25–29 October 2021

Relevant tasks in Nuclear Data

- Efficient information extraction from experimental measurements and associated publications
- UQ for all input of nuclear data evaluations including outlier detection, uncertainty estimate of model parameters & experimental data, etc.
- Validation of nuclear data libraries
- Generation/Improved accessibility of databases of sensitivities of integral responses to nuclear data and experimental data
- Trainings on how to apply ML methods in the nuclear data pipeline
- Consideration of physics constraints, e.g., through probability distributions, on nuclear theory model parameters (method developments)

Technical requirements

- Good databases
- Efficient use of computer simulations
- Use of emulators where necessary
- Documentation of APIs

State of the art

- Nuclear and atomic data evaluation:
  - Generalized Least Squares method still prevalent
  - Gaussian processes applied and show promise
  - Monte Carlo approaches (e.g., BMC) continue to improve
  - ML methods applied to determine parameters of physical models and fitting functions (e.g., R-Matrix)
  - First successful applications of neural networks (e.g., mass evaluations)

State of the art

- Validation:
  - ML approaches can identify incorrect nuclear data, problematic features of experiments and potential biases
  - Bayesian experimental design process (ML-driven approach to find optimal designs)

Limitations

IAEA

- Common Gaussian distribution assumption is often not justified (e.g., outliers)
- ML approaches require **high-quality data**, which is an obstacle at present
- Information in **databases very difficult to parse** automatically
- **Unbalanced UQ and quality control efforts** in different domains of data (weakest link determines overall quality)

Possible next steps

IAEA

- Continue **identifying important open questions/issues** that could benefit from the application of ML (our WG already made significant progress in this respect)
- **Enrich experimental databases** (e.g., EXFOR, ALADDIN) with well-documented API and rich meta data to facilitate the development and testing of ML approaches
- **Define diverse and broad test sets** with domain-specific data to validate ML models (relevant physics quantities) for comparison studies
- **Educational activities**, e.g., workshops, to teach ML methods, nuclear physics knowledge relevant for the application of ML
- **Crowd-sourcing campaigns** / open-science challenges to improve databases, models, etc. (e.g., hackathons, Kaggle as example)

Accelerating Progress – IAEA's role

- IAEA hosts prominent databases, e.g., EXFOR, RIPL, etc. so bring database maintainers, ML developers, and nuclear data users together to discuss how to **make databases more accessible** for ML studies.

- Draft a document laying out a **standard approach for data handling** (e.g., FAIR) and collaborative open-source practices with a focus on EXFOR and potentially others, considering input from nuclear data users, ML developers and database maintainers

- Organize meetings for **comparison exercises and educational workshops**

- Host repositories with **reference datasets** for domain-specific ML studies

Expected Outcomes

IAEA

- **Reference test sets** to validate ML models
- **Clear licensing** of data in databases and repositories
- **Roadmap** to future measurements and theory needs
- **Detection of issues** in data and models impeding scientific progress
- Significant speed-up of **evaluation and validation** work through the application of ML

IAEA
International Atomic Energy Agency

Thank you!