Mathematically Rigorous Verification & Validation of Scientific Machine Learning

Laura L. Pullum[1,2], Chad Steed[2], Sumit K. Jha[3], Arvind Ramanathan[2]

A successful program of **mathematically rigorous verification and validation** (V&V) **of scientific machine learning** (ML) will **enable** users and **decision-makers to justifiably trust this software** in terms of its fitness for use, reliability and robustness. This will help maximize the positive impact of scientific ML on DOE-mission and scientific/engineering applications and further the acceptance of, and advancement in, ML for DOE-mission and science/engineering applications.

Regardless of the software type, application or programming style, verification ("Did we build the software correctly?") and validation ("Did we build the right thing/software?") (V&V) comprise techniques and processes to ensure the resulting software meets the user's needs. Although V&V techniques and processes are well-established for typical deterministic software, they are either missing or inadequate for ML software. There is a fundamental mismatch between the importance of DOE-mission applications using scientific ML and the V&V that can currently be conducted.

Mathematical Research for Scientific ML V&V

The **key challenges** for mathematically rigorous V&V are: (i) There is typically a lack of a physics-based oracle against which to V&V ML results; (ii) The high-dimensionality of the hypothesis and problem spaces introduces risk in visual support for V&V; (iii) Nonlinear computations reduce or eliminate the usefulness of standard deterministic software V&V techniques; (iv) Probabilistic reasoning, with different computations and results from different runs of the same software, requires the ability to mathematically pose and evaluate "correctness" classes based on the hypothesis class and data space and (v) V&V of massively parallel implementations requires coupling abstractions and probabilistic specifications.

Suggested **new research directions** include:

***Validation*:** Extend the mathematical foundations of ML validation theory to define data validation (e.g., "Does the data comprise a valid representation of the problem space?") and valid boundaries for the ML problem itself. Addresses (i) and (iv); supports (v).

State-of-the-art (SOA) in Validation: ML validation has focused on cross-validation and information loss, and while necessary, this is insufficient in that it often disregards data validation and model (and parameter) selection. To date, little R&D has been conducted in mathematically rigorous validation for scientific ML.

***Visualization*:** Mathematically determine the information loss from abstractions required in high-dimension visualization and determine the bounds of information loss beyond which the visualization is inadequate for V&V support. Addresses (ii).

SOA in visualization (for ML V&V): Visualization has been quite beneficial in aiding the user/decision-maker's understanding of the input space and results of ML. The authors know of no research to date that addresses this research thread (information loss from abstraction in visualization).

***Nonlinear abstractions and adversarial attacks*:** a) Develop mathematical machinery for building novel, formal abstractions of scientific ML algorithms. A library of customized abstractions for different nonlinear operations may be required to perform static analysis of scientific ML. From the abstractions, adversarial attacks for verification can be developed. Addresses (iii). b) Massive parallelization employed in scientific ML requires the development of coupling abstractions like communicating sequential processes with probabilistic reasoning to design a notion of distributed Markov chains that can capture the specifications of massively parallel scientific ML software. Are specifications that enable programmers to

[1] Corresponding author: Dr. Laura L. Pullum, pullumll@ornl.gov, Oak Ridge National Laboratory (ORNL), 1 Bethel Valley Road, P.O. Box 2008, MS-6085, Oak Ridge, TN 37831-6085, (865) 574-4602
[2] Oak Ridge National Laboratory (ORNL)
[3] University of Central Florida

annotate scientific ML implementations with invariants and assertions that capture the expected behavior of the scientific algorithm useful for verification and automated runtime monitoring? Addresses (v)

SOA in abstraction and adversarial attacks. Adversarial attacks on high-dimension data sets have shown that empirical guarantees established using large benchmark data sets do not guarantee the correctness of ML algorithms on new or minimally perturbed data inputs [6, 7]. While adversarial synthesis has recently been conducted on images and videos interacting with deep learning algorithms [8], to the best of our knowledge, adversarial attacks on scientific data sets have not been investigated.

SOA in specifications. Existing work in specification languages such as the Java Modeling Language has focused more on control and less on hierarchical data sets critical to scientific ML.

Maturity, Uniqueness and Novelty

Maturity: Mathematical foundations of ML validation exist in ML theory and model selection and upon these the ML validation research thread can build. Visualization information loss will draw from existing information loss formulations, but will extend the application to loss due to ML abstraction. While fundamental research into static software analysis, distributed Markov chains and specification languages for runtime monitoring exists, there is little work on nonlinear abstractions for static analysis and probabilistic specifications for parallel ML. Initial experiments have been conducted in image and video classification [5, 6] and cyber-physical systems [6, 7].

Uniqueness: DOE-unique application domains include the smart grid and other cyber-physical systems, advanced/additive manufacturing modeling and scientific discovery. Relevant applications in which the authors have experienced the need for mathematically rigorous scientific ML V&V include Smart Grid applications [1, 2], predictive models for intelligence [3-5], cyber-physical systems [6, 7], data-driven models used in additive/advanced manufacturing[4], medical devices [8] and prognostic systems [9].

Novelty: The authors are not aware of other programs (e.g., at DARPA, NSF, etc.) that address the research needs identified herein. DARPA recently initiated a program in explainable artificial intelligence (AI) that, if successful, should help in understanding the how an AI application determines its output. However, the program has no results to date and it is unclear whether a mathematical approach will be taken.

References

1. IEEE, *IEEE Smart Grid Vision for Computing: 2030 and Beyond*, IEEE Computer Society Press, 2013.
2. Pullum, Laura L., et al. "Big Data Analytics, Machine Learning and Artificial Intelligence in the Smart Grid," IEEE Smart Grid white paper, online at the IEEE Resource Center for Smart Grid. 2017.
3. Pullum, L.L., "*Predictive Model*, v 1.0 Test Report," ORNL Technical Report for Department of Homeland Security S&T Office, 2016. [FOUO]
4. Pullum, L.L. and A. Ramanathan. Quantitative Approaches to Verify and Validate Anomaly Detection Algorithms. ORNL/LTR-2015/589. Oak Ridge, TN: Oak Ridge National Laboratory, 2015.
5. Ramanathan, A., L.L. Pullum, F. Hussain, D. Chakrabarty, and Sumit K. Jha. "Integrating symbolic and statistical methods for testing intelligent systems: Applications in machine learning and computer vision," 2016 Design, Automation and Test in Europe Conference and Exhibition (DATE), Dresden, pp. 786-791.
6. Raj, S., S.K. Jha, L.L. Pullum, and A. Ramanathan. Statistical Hypothesis Testing Using CNN Features for Synthesis of Adversarial Counterexamples to Human & Object Detection Vision Systems. ORNL TR, 6/1/17.
7. Raj, S., S.K. Jha, A. Ramanathan, and L.L. Pullum. "Work-in-progress: testing autonomous cyber-physical systems using fuzzing features from convolutional neural networks," 2017 Intl Conf on Embedded Software (EMSOFT), Seoul, 2017, pp. 1-2.
8. Pullum, L.L., and C. Symons, "Failure analysis of a complex learning framework incorporating multi-modal and semi-supervised learning," IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2011), 308-313, 2011.
9. Pullum, L.L., M. Darrah, S. Skias, K. Tso, and A. Tai, "Developing a data-driven prognostic system with limited system information," IEEE Intl Symp on High Assurance Systems Eng (HASE), Tampa, FL, 2004.

[4] Corresponding author is a working member (volunteer) of the ASME V&V 50 subgroup developing standards for the V&V of data-driven models in advanced manufacturing.