Discovering human-interpretable solution operators from realistic scientific data

Chris Earls
Center for Applied Mathematics
Cornell University

Scientific theories may be viewed as explanatory, predictive stories that describe the behavior of our physical world, and enable human technology. In so far as these theories are precise, they are expressed through a syntax founded on mathematics. Such predictive stories appeal to human minds, by construction. Indeed, the entire enterprise of science is human at its core; even having been identified as one of the greatest human accomplishments, by our Workshop's Organizing Committee.

With this view in mind, a profitable application of artificial intelligence to the scientific enterprise is one where human intelligence is tightly coupled with an artificial one. This talk will explore the role of solution operator learning to enable pioneering scientific discovery along frontiers that have defied our best efforts at scientific understanding until now; due to complex, nonlinear relations that have been obscured within high-dimensional data sets. Such physical contexts pose a challenge to the pattern recognition abilities of human minds; here artificial intelligence can play a crucial, complementary role in scientific discovery, but it must speak our language: mathematics.

Outlined now are my answers to the seven questions posed by our Workshop's Organizing Committee:

The original discovery problem they wanted to solve: Governing rate equations form the cornerstone of physics, and their compact solution operators[1–3] offer a wealth of physical insight. Discovering these from within sparse, noisy, high-dimensional cause-effect experimental data will yield mechanistic insights that lead to novel, testable scientific hypotheses: the grist of the scientific method.
How they formulated the problem in computational terms: This presentation introduces statistical learning and deep learning approaches for solution operator learning.
What data and knowledge they provided to their system: Our focus is on developing algorithms that work with realistic scientific data: sparse, noisy, and potentially biased.
How they represented the system's inputs and outputs: We discover our solution operators by collecting cause-effect data in the form of observed source-response behavior emanating from natural systems.
The space of candidate models that the system searched: The focus is on exploration within linear, compact operators, locally defined on some parameter space: nonlinearity may be considered through differential geometric arguments; leading to discovery of continuous, locally-linear solution operators.
What criteria it used to evaluate candidate models: Once discovered, these solution operators can be mined for mechanistic insight implied in underlying symmetries, invariants, and conserved quantities.
How they interpreted results that the system generated: Since the proposed systems are learning from data, using sophisticated mathematics, the discovered results offer physical insights on systems where such interpretations have eluded our best efforts, until now.

1. Praveen, H., Boulle, N. & Earls, C. Principled interpolation of Green's functions learned from data. (2022) doi:10.48550/ARXIV.2211.06299.
2. Boullé, N., Earls, C. J. & Townsend, A. Data-driven discovery of Green's functions with human-understandable deep learning. *Sci Rep* **12**, 4824 (2022).
3. Gin, C. R., Shea, D. E., Brunton, S. L. & Kutz, J. N. DeepGreen: deep learning of Green's functions for nonlinear boundary value problems. *Sci Rep* **11**, 21614 (2021).