1344
doi:10.1017/S1431927621005018

Microsc. Microanal. 27 (Suppl 1), 2021

Physics-guided machine learning: A new paradigm for scientific knoweldge discovery

Xiaowei Jia

University of Pittsburgh, Sewickley, Pennsylvania, United States

Process-based models of dynamical systems are often used to study engineering and environmental systems. Despite their extensive use, these models have several well-known limitations due to incomplete or inaccurate representations of the physical processes being modeled. Given rapid data growth due to advances in sensor technologies, there is a tremendous opportunity to systematically advance modeling in these domains by using machine learning (ML) methods. However, capturing this opportunity is contingent on a paradigm shift in data-intensive scientific discovery since the "black box" use of ML often leads to serious false discoveries in scientific applications. Because the hypothesis space of scientific applications is often complex and exponentially large, an uninformed data-driven search can easily select a highly complex model that is neither generalizable nor physically interpretable, resulting in the discovery of spurious relationships, predictors, and patterns. This problem becomes worse when there is a scarcity of labeled samples, which is quite common in science and engineering domains.

In this research, we will develop a novel methodology for combining process-based models with state-of-the-art machine learning models to leverage their complementary strengths. Our objective is to develop innovative process-guided deep learning models to better capture the dynamics in scientific systems and advance the understanding of underlying physical processes. Effective representation of physical processes will require development of novel abstractions and architectures that can simulate these processes that may be evolving and interacting at multiple scales. In addition, the optimization process to produce an ML model will have to consider not just accuracy (i.e., how well the output matches the observations) but also its correctness from a physical perspective (i.e., physical consistency).

Leveraging physics, both directly (e.g., by incorporating fundamental physical laws), and indirectly (e.g., as embodied in mechanistic models), is the key to address limitations of traditional ML models: inability to generalize to unseen scenarios and need for massive amounts of training data. Incorporation of relevant physical constraints, such as conservation of mass and energy, in the ML model helps ensure that the results produced will be physically consistent. At the same time, these constraints reduce the size of the hypothesis space, which means that less data is required for training. Violations of these physical laws in the output produced by ML models in future (previously unseen) scenarios can indicate that the models are no longer applicable to the new scenario, and thus the models become more trustworthy.

Specifically, we propose Process-Guided Recurrent Neural Network models (PGRNN) [1-3] as a general framework for modeling physical phenomena in many disciplines. As an alternative to both process-based and empirical models, PGRNN enriches the spatial and temporal modeling structure in the ML model by incorporating physical laws and generalizes the loss function to include a process-based penalty (based on energy conservation). Most importantly, the PGRNN model can be effectively trained even in absence of observed data due to its unique power to leverage the knowledge encoded in process-based models which were developed by domain scientists over many years.

This work makes a case that in real-world systems that are governed by physical processes, there is an opportunity to take advantage of fundamental physical principles to inform the search of a physically meaningful and accurate ML model. While we will illustrate this paradigm in the context of modeling water temperature, it has the potential to greatly advance the pace of discovery in a number of scientific and engineering disciplines where process-based models are used, e.g., power engineering, climate science, weather forecasting, materials science, and biomedicine.

References
[1] Jia, Xiaowei, Jared Willard, Anuj Karpatne, Jordan Read, Jacob Zwart, Michael Steinbach, and Vipin Kumar. "Physics guided RNNs for modeling dynamical systems: A case study in simulating lake temperature profiles." In Proceedings of the 2019 SIAM International Conference on Data Mining, pp. 558-566. Society for Industrial and Applied Mathematics, 2019.
[2] Read, Jordan S., Xiaowei Jia, Jared Willard, Alison P. Appling, Jacob A. Zwart, Samantha K. Oliver, Anuj Karpatne et al. "Process-guided deep learning predictions of lake water temperature." Water Resources Research (2019).
[3] Jia, Xiaowei, Anuj Karpatne, Jared Willard, Michael Steinbach, Jordan Read, Paul C. Hanson, Hilary A. Dugan, and Vipin Kumar. "Physics guided recurrent neural networks for modeling dynamical systems: Application to monitoring water temperature and quality in lakes." arXiv preprint arXiv:1810.02880 (2018).

https://doi.org/10.1017/S1431927621005018 Published online by Cambridge University Press