Optimal transfer learning strategies for predicting material properties

Sai Gautam Gopalakrishnan[1], Reshma Devi[1], Keith T. Butler[2]
[1]Department of Materials Engineering, Indian Institute of Science, Bengaluru 560012, India
[2]Department of Chemistry, University College London, London WC1E 6BT, United Kingdom

saigautamg@iisc.ac.in

Abstract

Materials science is a domain characterised by 'small' datasets (i.e., < 10,000 datapoints) of critical properties that govern performance of various applications and devices. For instance, there are no large, reliable datasets available for several key 'performance determining' metrics in energy applications, such as diffusivities in battery electrodes, carrier recombination rates in photovoltaics, and molecular adsorption energies for catalysis. On the other hand, there are reasonably 'large' datasets (> 100,000 datapoints) available on some properties, such as, bulk formation enthalpies, computed band structures, and crystal structures across wide chemical spaces. Thus, if key chemical, compositional, and structural trends can be captured in available large datasets and subsequently transferred (or re-learnt), it will enable the use of deep learning and graph based neural network models in smaller datasets as well. Hence, my talk will explore the utility of current transfer learning (TL) approaches that are available for computational materials science and identify optimal ways to employ TL-based strategies [1]. Specifically, TL involves training a neural network model on a larger dataset and subsequently retraining a fraction of the model on a smaller dataset. I will quantify the accuracy, transferability, and efficiency of TL models compared to models that have been trained from scratch. Finally, I will focus on TL models that can generalise over multi-properties during pre-training and can efficiently be re-trained on small datasets, which pave the way towards creating more general, foundational models, in the near future.

References

[1] Reshma Devi, Keith T. Butler, Gopalakrishnan Sai Gautam, npj Computational Materials, 10 (2024) 300.