World Academy of Science, Engineering and Technology
International Journal of Materials and Metallurgical Engineering
Vol:18, No:09, 2024

High-Fidelity Materials Screening with a Multi-Fidelity Graph Neural Network and Semi-Supervised Learning

Authors : Akeel A. Shah, Tong Zhang

Abstract : Computational approaches to learning the properties of materials are commonplace, motivated by the need to screen or design materials for a given application, e.g., semiconductors and energy storage. Experimental approaches can be both time consuming and costly. Unfortunately, computational approaches such as ab-initio electronic structure calculations and classical or ab-initio molecular dynamics are themselves can be too slow for the rapid evaluation of materials, often involving thousands to hundreds of thousands of candidates. Machine learning assisted approaches have been developed to overcome the time limitations of purely physics-based approaches. These approaches, on the other hand, require large volumes of data for training (hundreds of thousands on many standard data sets such as QM7b). This means that they are limited by how quickly such a large data set of physics-based simulations can be established. At high fidelity, such as configuration interaction, composite methods such as G4, and coupled cluster theory, gathering such a large data set can become infeasible, which can compromise the accuracy of the predictions - many applications require high accuracy, for example band structures and energy levels in semiconductor materials and the energetics of charge transfer in energy storage materials. In order to circumvent this problem, multi-fidelity approaches can be adopted, for example the Δ-ML method, which learns a high-fidelity output from a low-fidelity result such as Hartree-Fock or density functional theory (DFT). The general strategy is to learn a map between the low and high fidelity outputs, so that the high-fidelity output is obtained a simple sum of the physics-based low-fidelity and correction, Although this requires a low-fidelity calculation, it typically requires far fewer high-fidelity results to learn the correction map, and furthermore, the low-fidelity result, such as Hartree-Fock or semi-empirical ZINDO, is typically quick to obtain, For high-fidelity outputs the result can be an order of magnitude or more in speed up. In this work, a new multi-fidelity approach is developed, based on a graph convolutional network (GCN) combined with semi-supervised learning. The GCN allows for the material or molecule to be represented as a graph, which is known to improve accuracy, for example SchNet and MEGNET. The graph incorporates information regarding the numbers of, types and properties of atoms; the types of bonds; and bond angles. They key to the accuracy in multi-fidelity methods, however, is the incorporation of low-fidelity output to learn the high-fidelity equivalent, in this case by learning their difference. Semi-supervised learning is employed to allow for different numbers of low and high-fidelity training points, by using an additional GCN-based low-fidelity map to predict high fidelity outputs. It is shown on 4 different data sets that a significant (at least one order of magnitude) increase in accuracy is obtained, using one to two orders of magnitude fewer low and high fidelity training points. One of the data sets is developed in this work, pertaining to 1000 simulations of quinone molecules (up to 24 atoms) at 5 different levels of fidelity, furnishing the energy, dipole moment and HOMO/LUMO.

Keywords : .materials screening, computational materials, machine learning, multi-fidelity, graph convolutional network, semi-supervised learning
Conference Title : ICMMSE 2024 : International Conference on Metallurgy, Materials Science and Engineering
Conference Location : Paris, France
Conference Dates : September 16-17, 2024