Machine Learning to Investigate Material Properties: Application to Novel Solar Cell Materials

A. Gagliardi
Technische Universität München, Electrical and Computer Engineering,
München, Germany

The use of machine learning techniques in material science is rapidly evolving in recent years [1]. The dramatic success of Machine learning in many different fields highlights the statistical benefits of marrying highly non-linear and near-nonparametric models with large datasets, with efficient optimization algorithms running in distributed computing environments. Machine learning models project input data through several layers of nonlinearity and learn different levels of abstraction generating a link between the inputs and specific target outputs (or labels). In this talk two applications of Machine learning will be presented: one to predict charge mobility in organic semiconductor films and the other to screen lead-free perovskite optical bandgaps for photovoltaic applications.

Charge transport in organic semiconductors is extremely sensitive to structural changes on the molecular scale. This provides the opportunity to tune macroscopic quantities like the charge carrier mobility by manipulating the microscopic composition of the material. The connection between these macroscopic properties and the microscopic structure can be performed by simulation via a multiscale approach where different models are coupled in a hierarchical way to bridge the different spatial scales. However, the demanding electronic structure calculations on the molecular scale require certain approximations to allow for reasonable computation times. In the multiscale model presented here, charge transport in an amorphous organic thin film of pentacene is calculated following several steps: the molecular structure of the pentacene film is obtained by Molecular Dynamics simulations, the transfer integrals between molecules are calculated with Density Functional Theory (DFT) methods and passed to a kinetic Monte Carlo simulation to calculate charge carrier mobility. Since DFT simulations for every possible molecule orientation are not feasible, a machine learning approach to improve the efficiency in predicting the charge transfer integral is implemented [2]. One critical step to obtain a well trained, highly predictive model is to find a good representation for the molecular structure. With a better representation of the input data, the machine learning algorithm can predict the output data more accurately. Selecting the most suitable set of inputs requires insight into the underlying physics as well as the operation of the machine learning algorithm. Different geometric features like the distance between molecules, the orientation of molecules relative to each other, and the curvature of the molecules are used, however pure geometric features have some limits to predict transfer integrals accurately. As additional information, we use the Coulomb matrix as an input feature [3]. The Coulomb matrix contains the repulsion between all the atoms of a dimer. One advantage of this representation is that it is invariant to translations and rotations of the molecule.

There is growing evidence that inorganic lead-free mixed halide perovskites hold promise for resolving current issues in stability, toxicity, and efficiency of the standard perovskite. A mixed cation and anion strategy has enabled compositional engineering of bandgaps, enhanced cell stability, and extension of the available composition space for a diverse array of compounds [4]. Recent progress in all-inorganic perovskite solar cells (PSCs) has garnered excitement for their potential to address long term stability issues and tune performance [5]. However, major challenges remain in identifying good candidate compounds and charting out

material trends. One problem is that even the smallest composition spaces can be surprisingly complex: for example $MABI_3$ with Sn^{2+}/Pb^{2+} mixing at the B-site can display highly non-trivial bandgap trends [6] while X-site mixing of $CsPbX_3$ with I^-, Br^-, or Cl^- can produce anomalous alloy effects [7]. Making matters more difficult, single perovskite compositions alone can exhibit rich phase behaviour, complex sublattice ordering and lattice distortion. A comprehensive search for meaningful material-property relationships therefore requires a dense sampling of both the composition and configuration spaces, and quickly becomes intractable beyond consideration of only a few elements under strict constraints. To effectively engineer the properties of lead-free inorganic perovskites, a new approach is needed to draw meaningful inferences from limited data and accurately model complicated trends. Supervised machine learning can be applied to overcome these challenges. Here we show how our machine is capable of learning highly complex relationships using kernel ridge regression (KRR), yielding an effective model for the rapid prediction of electronic, geometric, and thermodynamic properties. This is achieved by shifting focus from the composition to a shared set of fundamental elemental properties, making the inputs to the algorithm both chemically intuitive and extremely versatile.

[1] K. T. Butler et. al., *Nature*, **2018**, 559, 547
[2] J. Lederer et. al., *Adv. Theory Simul.*, **2018**, *102*, 1800136
[3] K. Hansen et. al., *J. Chem. Theory Comput.*, **2013**, 9, 3404
[4] M. Saliba et al., *Science* **2016**, 354, 206-209
[5] Liang, J. et al. *Journal of the American Chemical Society* **2016**, 138, 15829-15832
[6] F. Hao et al. ,*Journal of the American Chemical Society* **2014**, 136, 8094-8099
[7] W.-J. Yin et al., *The Journal of Physical Chemistry Letters* **2014**, 5, 3625-3631