[L6] Applications of machine learning and artificial intelligence to designing chemicals and materials with the desired properties

Alexander Tropsha

Laboratory for Molecular Modeling, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, 27599, USA.

Historically, discovery of novel chemical entities has been driven by a laborious trial-and-error process. The growth of chemical and materials databases coupled with Big Data-enabled machine learning and artificial intelligence approaches offer the opportunity to transform this practice into data- and knowledge-driven rational design—accelerating the discovery of novel compounds with the desired properties. I will discuss applications of these approaches to the design of novel chemicals and materials.

Using data from the AFLOW repository of high-throughput ab-initio calculations for inorganic molecules, we have generated Quantitative Materials Structure-Property Relationship (QMSPR) models to predict critical material properties, metal/insulator classification, bulk modulus, Fermi energy, and band gap energy. To enable these calculations, we have developed novel materials descriptors such as universal property-labelled fragments (PLMF) [1]. We have established that the accuracy of predictions obtained with machine learning models matches that of GGA DFT functionals yet model development requires a minute fraction of computational time as compared to ab initio calculations. Notably, due to the representation of materials with PLMF, the QMSPR models are broadly applicable to virtually any stoichiometric inorganic materials. This representation also affords straightforward model interpretation in terms of simple heuristic design rules that could guide rational design of novel materials. As a proof-of-concept study we have employed a materials informatics approach to identify a novel photocathode material for dye-sensitized solar cells (DSSCs) [2]

In a separate study, we have developed a novel computational strategy based on deep and reinforcement learning techniques for de-novo design of molecules with desired properties. This strategy integrates two deep neural networks – generative and predictive – that are trained separately but employed jointly to generate novel chemical structures with the desired properties. Generative models are trained to produce chemically feasible SMILES, and predictive models are derived to forecast the desired compound properties. In the first phase of the method, generative and predictive models are trained separately with supervised learning algorithms. In the second phase, both models are trained jointly with reinforcement learning approach to bias newly generated chemical structures towards those with desired physical and biological properties. In the proof-of-concept study, we have employed this strategy to design chemical libraries biased toward compounds with either maximal, minimal, or specific ranges of physical properties, such as melting point and hydrophobicity, as well as to develop novel putative inhibitors of JAK2. This new approach can find general use for generating targeted chemical libraries optimized for a single desired property or multiple properties.

Bibliography:

[1] Isayev O, Oses C, Toher C, Gossett E, Curtarolo S., Tropsha A. Universal fragment descriptors for predicting properties of inorganic crystals. Nat Commun. 2017, 8, 15679. doi: 10.1038/ncomms15679
[2]. Moot, T., Isayev, O., R.W., McCullough, S.M., Zemaitis, M., Lopez, R., Cahoon, J.F., Tropsha, A. Material informatics driven design and experimental validation of lead titanate as an aqueous solar photocathode. Materials Discov., 2016, 6, 9-16.
[3]. Popova, M., Isayev, O., Tropsha, A. Deep Reinforcement Learning for De-Novo Drug Design. https://arxiv.org/abs/1711.10907.