Check for updates

Themed collection on Insightful Machine Learning for Physical Chemistry

Aurora E. Clark, (iD) *[a] Pavlo O. Dral, (iD) *[b] Isaac Tamblyn (iD) *[cd] and
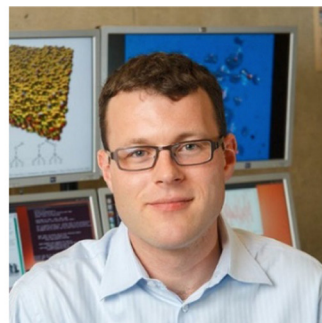Olexandr Isayev (iD) *[e]

Aurora E. Clark   Pavlo O. Dral   Isaac Tamblyn   Olexandr Isayev

Throughout the history of machine learning (ML) in Chemistry, there has been the dyad of *prediction* and *insight*. Within the last decade, the massive growth of ML in Chemistry has largely been driven by the availability of increasingly large datasets

[a] Department of Chemistry, University of Utah, Salt Lake City, UT 84112, USA.
E-mail: aurora.clark@utah.edu

[b] State Key Laboratory of Physical Chemistry of Solid Surfaces, College of Chemistry and Chemical Engineering, Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, and Innovation Laboratory for Sciences and Technologies of Energy Materials of Fujian Province (IKKEM), Xiamen University, Xiamen, Fujian 361005, China. E-mail: dral@xmu.edu.cn; Web: https://dr-dral.com

[c] Department of Physics, University of Ottawa, Canada. E-mail: isaac.tamblyn@uottawa.ca

[d] Vector Institute for Artificial Intelligence, Toronto, ON M5G 1M1, Canada

[e] Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, USA. E-mail: olexandr@olexandrisayev.com

that have greatly improved the predictive capabilities of ML. While property prediction has historically played a major role in Chemical machine learning studies, understanding the "how" and the "why" is an ongoing challenge addressed within this themed collection on *Insightful Machine Learning for Physical Chemistry*. This collection reflects the growing trend to explore new approaches that provide insight within various stages of machine learning, to develop new conceptual models and improve fundamental chemical theories. The reported rigorous and creative workflows are designed to teach us how the fundamental physics of atoms and electrons manifest themselves in molecular physicochemical phenomena and how this information traverses scale to condensed-matter behavior.

Predictions with fundamental chemical theories, particularly quantum chemistry, depend on a number of factors: the

approximate Hamiltonian employed, basis-set dependence, and the consideration of environmental effects (gas phase, cluster or embedded continuum). These theories are currently overhauled by ML to improve both their accuracy and speed, as is demonstrated by several studies in this issue that use Δ-learning and transfer-learning techniques. Such studies by themselves provide a unique insight into the importance of different molecular features in property prediction. A different mindset is adopted in several studies that explore the interrelation between quantum chemical methods and machine learning. As noted in the work of Kulik *et al.* in **https://doi.org/10.1039/D3CP00258F**, one might anticipate significant correlations of predicted molecular properties of transition-metal complexes amongst similar density functional approximations or families of functionals; yet, many features are relatively insensitive to functional-dependent errors,

This journal is © the Owner Societies 2023

Phys. Chem. Chem. Phys., 2023, **25**, 22563–22564 | **22563**

which supports an expanded view of virtual high-throughput screening using ML with density functional theory. The Perspective by Manzhos *et al.* in **https://doi.org/10.1039/D2CP04155C** shows that we might as well draw parallels between quantum chemistry and ML. The authors demonstrate how popular ML approaches and basis set expansions used in quantum chemical methods are interrelated, which may be useful to explore the limitations of both types of models for nonlinear, high-dimensional chemical problems.

Simulating complex environments and extended systems is another challenge tackled in the themed collection, where ML methods have been developed to predict solubilities and green solvents, along with redox potentials and optical absorption in solvents, to perform (QM)ML/MM molecular dynamics in the condensed phase, and study thermal transport across interfaces. Other studies in this collection have investigated how ML potentials can be built for larger systems based on smaller fragments.

Studying the reactivity of molecules within increasingly complex environments *via* new methods to calculate potential-energy landscapes (as in learned potentials of inter-particle interactions), through analysis and exploration of energy landscapes, or by connecting reactive sequences (as in reaction networks and kinetic modeling) has emerged as a significant topic. As illustrated in several works within this issue, reduced dimensional concepts associated with reactivity can bias our interpretation and understanding of the dynamic evolution of a chemical system. Yet this also provides an opportunity for learning how dimensionality reduction through eigende-composition, compression, clustering and other methods depend upon the sampling of input data across different dimensions and influence the resulting information content and interpretation. The work of Deng and coworkers in **https://doi.org/10.1039/D2CP05083H** offers interpretable Bayesian Chemical Reaction Neural Networks to incorporate and quantify uncertainty for competitive reaction pathways

within given confidence intervals; these are based upon probabilistic distributions of chemical concentrations and physical parameters from the Arrhenius law and stoichiometric coefficients within the reaction network. Here, optimization of different parameters through the lens of their probabilistic distributions is a fundamental step toward understanding uncertainty quantification. This study highlights another important facet of this collection, *i.e.*, that incorporating uncertainty is highly beneficial for our understanding of reactivity and properties.

As demonstrated through the works in this themed collection, there are a myriad of ways by which we can incorporate interpretation and insight as design features of machine learning workflows. As guest editors, we envision a future where the dyad of prediction and insight in ML is treated routinely on equal footing to accelerate discovery and innovation, while at the same time providing a basis to fundamentally improve chemical theories. Ultimately, this should be a feedback loop within the chemical enterprise.