MRS Communications 2019 • Volume 9, Issue 3

ARTIFICIAL INTELLIGENCE SPECIAL ISSUE: PROSPECTIVES

Symbolic regression in materials science

Yiqun Wang, Nicholas Wagner, and James Rondinelli, Northwestern University, USA

The authors showcase the potential of symbolic regression as an analytic method for use in materials research. First, they briefly describe the current state-of-the-art method, genetic programming-based symbolic regression (GPSR), and recent advances in symbolic regression techniques. Next, they discuss industrial applications of symbolic regression and its potential applications in materials science. They then present two GPSR use-cases: formulating a transformation kinetics law and showing the learning scheme discovers the well-known Johnson–Mehl–Avrami–Kolmogorov (JMAK) form, and learning the Landau free energy functional form for the displacive tilt transition in perovskite $LaNiO_3$. Finally, they propose that symbolic regression techniques should be considered by materials scientists as an alternative to other machine-learning-based regression models for learning from data. DOI.org/10.1557/mrc.2019.85

Materials science in the AI age: High-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics

Rama K. Vasudevan, Oak Ridge National Laboratory, USA; Kamal Choudhary, National Institute of Standards and Technology, USA; Apurva Mehta, SLAC National Accelerator Laboratory, USA; Ryan Smith, Gilad Kusne, and Francesca Tavazza, National Institute of Standards and Technology, USA; Lukas Vlcek, Maxim Ziatdinov, and Sergei V. Kalinin, Oak Ridge National Laboratory, USA; and Jason Hattrick-Simpers, National Institute of Standards and Technology, USA

The use of statistical/machine learning approaches to materials science is experiencing explosive growth. Here, the authors review recent work focusing on generation and application of libraries from both experiment and theoretical tools. The library data enables classical correlative machine learning, and also opens the pathway for exploration of underlying causative physical behaviors. The authors highlight key advances facilitated by this approach, and illustrate how modeling, macroscopic experiments, and imaging can be combined to accelerate understanding and development of new materials systems. These developments point toward a data-driven future wherein knowledge can be aggregated and synthesized, accelerating the advancement of materials science. DOI.org/10.1557/mrc.2019.95

Embedding domain knowledge for machine learning of complex material systems

Christopher Childs and Newell Washburn, Carnegie Mellon University, USA

Machine learning has revolutionized disciplines within materials science that have been able to generate sufficiently large datasets to utilize algorithms based on statistical inference, but for many important classes of materials the datasets remain small. However, a rapidly growing number of approaches to embedding domain knowledge of materials systems are reducing data requirements and allowing broader applications of machine learning. Furthermore, these hybrid approaches improve the interpretability of the predictions, allowing for greater physical insight into the factors that determine material properties. This review introduces a number of these strategies, providing examples of how they were implemented in machine learning algorithms and discussing the materials systems to which they were applied. DOI.org/10.1557/mrc.2019.90

Deep materials informatics: Applications of deep learning in materials science

Ankit Agrawal and Alok Choudhary, Northwestern University, USA

The growing application of data-driven analytics in materials science has led to the rise of materials informatics. Within the arena of data analytics, deep learning has emerged as a game-changing technique in the last few years, enabling numerous real-world applications, such as self-driving cars. In this article, the authors describe an overview of deep learning, its advantages, challenges, and recent applications on different types of materials data. The increasing availability of materials databases and big data in general, along with groundbreaking advances in deep learning offers a lot of promise to accelerate the discovery, design, and deployment of next-generation materials. DOI.org/10.1557/mrc.2019.73

ARTIFICIAL INTELLIGENCE SPECIAL ISSUE: RESEARCH LETTERS

Experiment Specification, Capture and Laboratory Automation Technology (ESCALATE): A software pipeline for automated chemical experimentation and data management

Ian Pendleton and Gary Cattabriga, Haverford College, USA; Zhi Li, Lawrence Berkeley National Laboratory, USA; Mansoor Ani Najeeb, Sorelle Friedler, and Alexander Norquist, Haverford College, USA; Emory Chan, Lawrence Berkeley National Laboratory, USA; and Joshua Schrier, Fordham University, USA

Applying artificial intelligence to materials research requires abundant curated experimental data and the ability for algorithms to request new experiments. ESCALATE (Experiment Specification, Capture and Laboratory Automation Technology)—an ontological framework and open-source software package—solves this problem by providing an abstraction layer for human- and machine-readable experiment specification, comprehensive and extensible (meta-) data capture, and structured data reporting. ESCALATE simplifies the initial data collection process, and its reporting and experiment generation mechanisms simplify machine learning integration. An initial ESCALATE implementation for metal halide perovskite crystallization was used to perform 55 rounds of algorithmically controlled experiment plans, capturing 4336 individual experiments. DOI.org/10.1557/mrc.2019.72

Active-learning and materials design: The example of high glass-transition temperature polymers

Chiho Kim, Anand Chandrasekaran, Anurag Jha, and Rampi Ramprasad, Georgia Institute of Technology, USA

Machine-learning (ML) approaches have proven to be of great utility in modern materials innovation pipelines. Generally, ML models are trained on predetermined past data and then used to make predictions for new test cases. Active-learning, however, is a paradigm in which ML models can direct the learning process itself through providing dynamic suggestions/queries for the "next-best experiment." In this work, the authors demonstrate how an active-learning framework

https://doi.org/10.1557/mrs.2019.244 Published online by Cambridge University Press

can aid in the discovery of polymers possessing high glass-transition temperatures (*T*g). Starting from an initial small dataset of polymer *T*g measurements, the authors use Gaussian process regression in conjunction with an active-learning framework to iteratively add *T*g measurements of candidate polymers to the training dataset. The active-learning framework employs one of three decision making strategies (exploitation, exploration, or balanced exploitation/exploration) for selection of the "next-best experiment." The active-learning workflow terminates once 10 polymers possessing a *T*g greater than a certain threshold temperature are selected. The authors statistically benchmark the performance of the aforementioned three strategies (against a random selection approach) with respect to the discovery of high-*T*g polymers for this particular demonstrative materials design challenge. DOI.org/10.1557/mrc.2019.78

Artificial neural network correction for density-functional tight-binding molecular dynamics simulations

Junmian Zhu, Grinnell College, USA; Van Quan Vuong, The University of Tennessee, Knoxville, USA; Bobby Sumpter, Oak Ridge National Laboratory, USA; and Stephan Irle, The University of Tennessee, Knoxville, and Oak Ridge National Laboratory, USA

The authors developed a Behler–Parrinello-type neural network (NN) to improve the density-functional tight-binding (DFTB) energy and force prediction. The Δ-machine learning approach was adopted and the NN was designed to predict the energy differences between the density functional theory (DFT) quantum chemical potential and DFTB for a given molecular structure. Most notably, the DFTB-NN method is capable of improving the energetics of intramolecular hydrogen bonds and torsional potentials without modifying the framework of DFTB itself. This improvement enables considerably larger simulations of complex chemical systems that currently could not easily have been accomplished using DFT or higher level *ab initio* quantum chemistry methods alone. DOI.org/10.1557/mrc.2019.80

Machine learning prediction of accurate atomization energies of organic molecules from low-fidelity quantum chemical calculations

Logan Ward, Argonne National Laboratory, and The University of Chicago, USA; Ben Blaiszik, Argonne National Laboratory, and Globus, USA; Ian Foster, Argonne National Laboratory, The University of Chicago, and Globus, USA; Rajeev S. Assary, Argonne National Laboratory, USA; Badri Narayanan, Argonne National Laboratory, and University of Louisville, USA; and Larry Curtiss, Argonne National Laboratory, USA

Recent studies illustrate how machine learning (ML) can be used to bypass a core challenge of molecular modeling: the tradeoff between accuracy and computational cost. The authors assess multiple ML approaches for predicting the atomization energy of organic molecules. Their resulting models learn the difference between low-fidelity, B3LYP, and high-accuracy, G4MP2, atomization energies, and predict the G4MP2 atomization energy to 0.005 eV (mean absolute error) for molecules with less than 9 heavy atoms (training set 117232, test set 13026) and 0.012 eV for a small set of 66 molecules with between 10 and 14 heavy atoms. Their best models, which have accuracy/speed tradeoffs, enable the efficient prediction of G4MP2-level energies for large molecules and are available through a web interface. DOI.org/10.1557/mrc.2019.107

Robocrystallographer: Automated crystal structure text descriptions and analysis

Alex Ganose and Anubhav Jain, Lawrence Berkeley National Laboratory, USA

Our ability to describe crystal structure features is of crucial importance when attempting to understand structure–property relationships in the solid state. In this paper, the authors introduce robocrystallographer, an open-source toolkit for analyzing crystal structures. This package combines new and existing open-source analysis tools to provide structural information, including the local coordination and polyhedral type, polyhedral connectivity, octahedral tilt angles, component-dimensionality, and molecule-within-crystal and fuzzy prototype identification. Using this information, robocrystallographer can generate text-based descriptions of crystal structures that resemble descriptions written by human crystallographers. The authors use robocrystallographer to investigate the dimensionalities of all compounds in the Materials Project database and highlight its potential in machine learning studies. DOI.org/10.1557/mrc.2019.94

Prediction of new iodine-containing apatites using machine learning and density functional theory

Timothy Hartnett, Mukil Ayyasamy, and Prasanna V. Balachandran, University of Virginia, USA

The authors develop a computational approach that integrates machine learning (ML) and density functional theory (DFT) with experimental data to predict formable and thermodynamically stable iodine-containing apatites. This is an important problem because radioactive iodine is toxic and capturing it in solid waste forms has implications in remediation treatments. They train ML models using 336 compositions and screen 54 iodine-containing compounds in apatite stoichiometry. ML models predict 18 as formable and 24 as non-formable in the apatite structure; 12 compounds were identified to be uncertain. DFT convex hull predicted two to be thermodynamically stable, one as metastable, and nine as unstable. DOI.org/10.1557/mrc.2019.103

PROSPECTIVE

A perspective on triplet fusion upconversion: Triplet sensitizers beyond quantum dots

Zachary A. VanOrman, Alexander S. Bieber, Sarah Wieghold, and Lea Nienhaus, Florida State University, USA

The processes of singlet fission and triplet fusion could allow state-of-the-art photovoltaic devices to surpass the Shockley–Queisser limit by optimizing the utilized solar spectrum by reducing thermal relaxation and inaccessible sub-bandgap photons, respectively. Both demand precise control of the spin-triplet state population, which, requires a sensitizer to populate the triplet state of an acceptor molecule. In this perspective, the authors highlight the established field of sensitized upconversion and further examine alternative triplet sensitization routes, including the possibility of bulk solid-state semiconductors as triplet sensitizers, which provide a new avenue for charge transfer based triplet sensitization rather than excitonic triplet energy transfer. DOI.org/10.1557/mrc.2019.115

Visit mrs.org/mrc to view additional Prospectives Articles, Commentaries, and Research Letters.

https://doi.org/10.1557/mrs.2019.244 Published online by Cambridge University Press