Editorial: Machine Learning and Data Mining in Materials Science

Huber, Norbert; Kalidindi, Surya R.; Klusemann, Benjamin; Cyron, Christian J.

frontiers
in Materials

Check for updates

Editorial: Machine Learning and Data Mining in Materials Science

Norbert Huber [1,2]*, Surya R. Kalidindi [3], Benjamin Klusemann [1,4] and Christian J. Cyron [1,5]

[1] Institute of Materials Research, Materials Mechanics, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany, [2] Institute of Materials Physics and Technology, Hamburg University of Technology, Hamburg, Germany, [3] School of Mechanical Engineering and School of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, United States, [4] Institute of Product and Process Innovation, Leuphana University of Lüneburg, Lüneburg, Germany, [5] Institute of Continuum and Materials Mechanics, Hamburg University of Technology, Hamburg, Germany

Editorial on the Research Topic

Machine Learning and Data Mining in Materials Science

The development of new materials, incorporation of new functionalities, and even the description of well-studied materials strongly depends on the capability of individuals to deduce complex structure-property relationships. A significant challenge in this field remains the "curse of dimensionality". Even for the characterization of moderately complex materials, often a considerable number of parameters is required to characterize their composition and microstructure (or also processing conditions) uniquely. Modeling of materials is thus facing the challenge of high-dimensional parameter spaces, where numerous parameter combinations have to be sampled and studied thoroughly. Relying thereby on experiments is typically prohibitively expensive, given the often high-dimensional parameter space of interest. Thus, the combination of experimental and computational approaches is receiving increasing attention. The complex interdependencies in the resulting data sets can be studied using machine-learning approaches. Artificial neural networks and data-driven approaches can significantly help to identify, approximate, and visualize structure-property relationships of interest. This way, they can accelerate our understanding and effective utilization of complex hierarchical materials.

This Research Topic is a compilation of contributions on current ideas and novel concepts for the advancement of machine learning, data mining, and data driven-approaches in the context of the design of materials and materials processing. This includes general methods as well as their application to decoding the complex relationships along the chain composition—processing—structure—mechanical properties.

The review article by Bock et al. provides an overview on the state of art about machine learning and statistical learning approaches in the field of continuum materials mechanics. Furthermore, works on experiment- and simulation-based data mining in combination with machine leaning tools are presented. The reviewed papers are categorized as descriptive, predictive, or prescriptive depending on whether they aim at identification, prediction, or even optimization of essential characteristics. The potential of utilizing machine learning in materials science to empower significant acceleration of knowledge generation is highlighted. The other review article within

this collection by Talapatra et al. discusses the need and challenges of optimal experimental setups as a key factor for accelerating the discovery of materials. The authors review the most important challenges and opportunities connected with the concept of optimal experiment design and present successful examples that have led to materials discovery via this concept.

Advances of machine learning and data mining methods are addressed in particular in three articles of this special issue. Fritzen et al. developed a multi-fidelity surrogate model allowing for an adaptive on-the-fly switching between different surrogate models for a concurrent two-scale simulation. The first surrogate model is based on reduced order modeling, where the second one represents an artificial neural network (ANN). The methodology provides a suitable basis for the generalization of the applied machine learning techniques for different applications. Aydin et al. show how the bottleneck of computational data generation can be widened by an effective combination of simulations with different accuracy and computational cost. A cheap low-fidelity computational model is used to start the training of the ANN and then gradually switches to higher-fidelity training data as the training of the ANN progresses. This multi-fidelity strategy can reduce the total computational cost by a half up to one order of magnitude. González et al. emphasize how to enhance suitable physical models by available experimental data. Rather than substituting physical models by data, the authors are using the data to correct and enhance the physical law/model of interest, ensuring thermodynamic consistency. Rather than creating a purely data-driven model, the proposed technique represents an appealing alternative for machine learning of models from data.

Characteristics of the lower scales often significantly influence or dominate the macroscopic behavior of materials, making an appropriate characterization of the lower scales indispensable. Unfortunately, common (crystal) structure identification techniques can often not be applied to describe the structure of individual atoms in grain boundaries (GBs) sufficiently. To address this problem, Snow et al. used a form of Common Neighbor Analysis for the identification and characterization of arbitrary atomic structures found around GBs. The resulting structure descriptors are used as input to machine learning algorithms, here PCA with linear regression, for the development of atomic structure-property models for GBs. In the same spirit, Homer et al. developed a new structural representation, called the scattering transform, for characterization of GBs. This approach uses wavelet-based convolutional neural networks to characterize grain boundaries. The learning results are compared to a SOAP (smooth overlap of atomic positions) based representation, which reveals some benefits on the scattering transform, e.g., learning well on larger datasets and providing physically interpretable information. At the microscale, Steinberger et al. used a machine learning based approach for classification of coarse-grained dislocation microstructures. As potential machine learning features, the dislocation microstructure is described via different dislocation density field variables. It is shown that the accuracy of machine learning models varies with different sets of microstructure features and spatial discretization. This can also be used as an indicator for testing the ability of a coarse-grained model to capture the underlying mechanisms accurately. At the macroscale, Furat et al. present

various applications for segmentation of tomographic imaging data by combining machine learning methods and conventional image processing techniques. They demonstrate the applicability of their approach using the example of grain-wise segmentation of time-resolved CT data obtained in between Ostwald ripening steps of an AlCu specimen. Richert et al. investigated algorithms used for the measurement of complex 3D microstructures with respect to over- and underestimation of the thickness of curved features, which can lead to a significant error in the prediction of mechanical properties. Here, artificial neural networks are applied for reconstruction of the true geometry from the image processing data within voxel resolution.

In terms of materials modeling along the process-property-structure-performance chain, Würger et al. successfully used a combination of experiments, machine learning, data mining, density functional theory, and molecular dynamic calculations to determine property-structure relationships in magnesium alloys with respect to corrosion. Corrosion inhibition properties of still untested molecules are estimated and a relationship between corrosion inhibition efficiency and corresponding molecular structure of magnesium corrosion inhibitors is established. Castillo and Kalidindi present a two-step Bayesian framework for the estimation of the intrinsic single crystal elastic stiffness parameters from the measurements of spherical indentation stress-strain responses in multiple individual grains of a polycrystalline sample, whose crystal lattice orientations have been measured using electron back-scattered diffraction technique. It is shown that the introduction of a Bayesian framework can greatly reduce the number of simulations necessary to establish this function. The novel framework is presented and demonstrated for a cubic polycrystalline Fe-3%Si sample and a hexagonal polycrystalline pure titanium sample. In the approach by Reimann et al., the macroscopic material behavior is described via a trained machine learning algorithm based on micromechanical simulations, i.e., uniaxial loading of representative volume elements of the microstructure of interest. In this regard, the trained algorithm can be interpreted as a macroscopic constitutive relation. The approach is illustrated for damage modeling as well as microstructure design that lead to targeted mechanical properties.

Menon et al. present a general hierarchical machine learning (HML) model for predicting the stress-at-break, strain-at-break, and Tan δ for thermoplastic and thermoset polyurethanes. The algorithm was trained on a library of 18 polymers. HML reduces data requirements through robust embedding of domain knowledge and surrogate data in a middle layer that bridges input variables (composition) and output responses (mechanical properties). The HML predictions are shown to be more accurate than those from a random forest model directly relating composition and properties, suggesting that embedding domain knowledge provides significant advantages in predicting the properties of complex material systems based on small datasets. Huber addresses a number of fundamental questions regarding the topological description of materials characterized by a highly porous three-dimensional structure. Via data mining, the interdependencies of topological parameters and relationships between topological parameters with mechanical properties are discovered. The determination of the average coordination

number turned out to be a difficult problem, which is solved by artificial neural networks by reconstructing the information on low-coordinated junctions that are not detectable from a common structure analysis.

The reviews and original articles compiled in this Research Topic give a taste of the potential of coupling approaches from materials science, modeling, and simulation with data mining and machine learning. This offers exciting perspectives for solving challenging problems, such as decoding and computational modeling of complex structure-process-property relationships, replacement of computationally demanding submodels in multiscale simulations, or classification and interpretation of imaging data.

AUTHOR CONTRIBUTIONS

This Editorial was jointly written by all authors who also served as Guest Editors for the Research Topic.

FUNDING

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.