Preface to the special issue on machine learning and data-driven design of materials issue in computational materials science

Taylor D. Sparks

Materials Science & Engineering Department, University of Utah, Salt Lake City, UT, 84112, United States

Anton O. Oliynyk

Chemistry & Biochemistry Department, Manhattan City College, Riverdale, NY, 10417, United States

Vitaliy Romaka

Department of Applied Material & Materials Processing, Lviv Polytechnic, Lviv, 79000, Ukraine

The discovery of materials has always been the cornerstone of civilizations. Empires have risen and fallen as the new materials were incorporated in the technologies of their time: Stone Age, Bronze age, Iron Age, Porcelain Age (Middle Ages), Steel Age (Industrial Era). It is strongly believed that the new Age (1950 C.E. - present) should be called Silicon Age. Not only Silicon has given us computers, but also a new revolutionary way to target novel materials with the guidance of computers: discovery with machine learning and data-driven approaches. Material discovery has historically been an incremental process. Past methods involved systematic study, literature-driven attempts, or serendipitous discovery. Unfortunately, these approaches are costly, time-consuming, and labor-intensive. A new paradigm was needed. In 2011 the Materials Genome Initiative (MGI) was issued "to discover, develop, and deploy new materials twice as fast at a fraction of the cost." The premise of the MGI was that combining experiment with computation and big data would accelerate materials discovery. The search for new materials could move beyond the stage of undirected, random attempts and instead rely on a systematic and semi-rational methods for materials discovery.

We have already seen successful demonstrations of machine-learning models trained on large data resulting in the prediction of novel topological, thermoelectric, superconductor, magnetic, phosphor, memory shape alloy, superhard materials and much more! These models rely on high-quality training data, chemical and structural descriptors, algorithms, cross validation, and screening. Machine learning often allows data scientists to work collaboratively with domain experts during data aggregation, data curation, and descriptor development. Computational materials science can provide novel training data and custom chemical and structural descriptors. Traditional approaches such as molecular dynamics, density functional theory, and phase field calculations can be thought of as complimentary tools rather than competing techniques for machine learning.

The manuscripts submitted to this special issue span a wide variety of topics related to materials science including graph representations of structure, attention-based learning, alloy development, diffusion, polymers, electronic properties, structure prediction, characterization, extrapolation beyond the training data set, additive manufacturing, kinetics, adaptive design and optimization. The guest editors would like to take this opportunity to express their gratitude to the many authors who made this special issue possible and we also extend our gratitude to the reviewers who put in time and effort in the review process.