Knowledge-integrated machine learning for materials: lessons from gameplaying and robotics
Executive Summary by Jacob Christensen, MSE 498, 7/7/2023

Introduction

Machine learning (ML), techniques where computers learn and grow from experience, is an area of great potential for materials research. However, bridging the gap between using ML to solve extensive engineering problems and the single-scope applications of today is a multi-faceted problem. Tasks such as materials selection and synthesis have seen improvements through ML assistance, but the interconnectivity and self-sustainment required to handle these problems independent from human involvement is yet to be realized. In this review, ML advances from two other fields, robotics and gameplaying, are used as examples for methods applicable to materials science. Three levels of implementation are identified and discussed. First, toolset integration, where ML is used to bolster experimental and simulation toolsets. Second, workflow integration, where ML is expanded to several steps for a larger workflow. Third, knowledge integration, where learnings from ML studies are autonomously pooled into databases.

Machine Learning Levels in Materials Research

Well-established ML algorithms have been introduced into materials research with techniques such as weighting and interpolation to account for sparse and biased data. Despite this, ML toolsets have often required significant human involvement to handle and process data, leaving the general laboratory workflow mostly unaltered. Therefore, an important task is to construct an interface between algorithms and instrumentation to gather, process, and interpret data without human intervention.

After single tool optimizations are complete, workflows can be developed to address different steps of the scientific process. For example, a suite of ML algorithms could search for new, stable materials for solar cells by funneling down possibilities based on desired properties and then rapidly characterizing the remaining materials. This could improve and grow materials databases through autonomous surveying. Picture an imaging lab which can perform rapid optical measurements and identify properties on its own, increasing throughput of simple characterization. However, adaptability to new tasks and analysis flexibility will remain areas where artificial intelligence (AI) like this can always be improved.

Fully realizing the potential of ML workflows requires dissemination of knowledge from one study to the next. This ensures that the efficiency and power of algorithms is maximized. Studies have been conducted to identify promising energy materials through ML by scanning millions of manuscript abstracts and finding patterns in desired properties. Collaboration, both direct and indirect like the above, is integral to increasing ML presence in materials research.

Machine Learning in Gameplaying and Robotics

The first implementations of AI in gaming included full solutions for every outcome. This becomes unmanageable quickly as games become more complex. Therefore, hand-crafted rulesets to govern how the algorithm should find optimized techniques were developed. But still, as the scope of the game in question increases, rulesets must follow, either becoming unreasonable to develop or too restrictive as they inevitably fall behind. Machine learning provided a remedy to this. For example, reinforcement ML algorithms allow AI to undergo training sessions in which desirable behavior and outcomes are rewarded. Training capabilities can be bolstered through manually demonstrating expert techniques. Now, an AI can adapt to new rulesets on its own, such as a varying board size in the game Go. Hybrid models may also be used for a balance between adaptability and computational time, combining the strengths of both model-based and learning-based systems.

Robotics have been a prime area to apply AI and ML. Abilities such as perception, sensory interpretation, and more complex tasks like connecting autonomous reasoning to appropriate action/response are enabled by ML. Multiple AI agents may be combined on a problem to interface both with the environment and each other, improving the learning process through cross-referencing and task breakdown. Allowing agents to interact with one another ensures each agent works with up-to-date information on the task, but incurs serious costs in complexity of the operation. When agents are set to ignore one another, complexity and time costs decrease, but this may undermine the guarantee of agreement among the agents' solutions.

Lessons from Gameplaying and Robotics

Within toolset integration, materials research can work to implement the capabilities of robots to produce output without necessarily copying human workflows. Research laboratories can either fully automate themselves for high throughput and consistency, or identify key points in the workflow that can be automated for maximum cost/benefit. Robotics can also assist in the transition from lab to manufacturing, a common end goal of materials research.

Within workflow integration, crafting well-defined targets helps communities focus on improving databases to strengthen the AI learning process. Compared to the robotics and gameplaying scenes, materials databases are lacking in numbers and are plagued with bias. Efforts must be made towards quantifying bias and setting standards for better interoperability. Open-source instrumental design, inspired from the robotics community, can assist in generating standardized data. Gameplaying demonstrated the strengths of hybrid ML designs which combine ruleset models and adaptable learning. One example of this in materials research so far is the improvement of solar cell manufacturing through constraining the search space with prior expert observations of materials quality and tool capabilities.

Knowledge integration is the most difficult level to achieve, characterized by three challenges. The first challenge is creating a connection between ML algorithms and knowledge pools which allow for both input and output in either direction. The second challenge is to make the knowledge bases more accessible among different studies. The third challenge is to develop ML algorithms that can effectively reuse pooled knowledge. Gameplaying demonstrates this with algorithms that work together when trained on an instance of a game, transferring that knowledge from one game to another without forgetting. To this date, materials research has primarily built models which specialize into niche studies, providing high accuracy but low generalizability. An effort should be made to increase generalizability in order to facilitate discovery, invention, and hypothesis generation among AI.

Conclusion and Outlook

The promise of more efficient, higher throughput research should press forward efforts to implement the ML advancements seen in robotics and gameplaying into materials science. Important lessons to learn include the benefits of combining human knowledge with learning algorithms, and the importance of developing community based goals. So far, materials research has shown progress in both toolset and workflow integration. Going forward, this field should strive to lead the development of knowledge integration, since a network of standardized databases can help make breakthroughs in device performance. Areas of creativity once thought exclusive to humans can be unlocked through knowledge-based ML. Now, it falls on humans to conduct extensive research into the ethical use of these accelerated and unsupervised AI, evaluate legality and misuse, and ensure products are safe for human health and the environment.