The combination of big data and AI is being called the "fourth industrial revolution," and its applications in materials science have soared in the past decade. The AI subfield of machine learning is already aiding the discovery of new materials.

Artificial intelligence is aiding the search for energy materials

By **Prachi Patel**
Feature Editor: **Shyue Ping Ong**

Early on, there was the incandescent light bulb. Now, buildings and streets are lit by light-emitting diodes (LEDs) that use less than 25% of the energy and last 25 times longer.

The LED is the result of painstaking years spent searching for the right semiconducting compounds and finessing their microstructures. And it's just one example of how materials science is pivoting the transition to a future where everyone has access to sustainable, affordable energy.

Satiating the needs and wants of an increasingly wealthy, digitized world will require new ways to convert and use energy. Materials are the foundation of those advances. As commercial silicon solar cells reach their theoretical limit, complementing them with efficient, affordable perovskite solar cells could help photovoltaic modules generate more power. The search is intensifying for advanced batteries that pack more energy than lithium-ion devices for lower cost and do not use scarce, expensive metals such as cobalt. Researchers are also investigating better thermoelectric materials, catalysts for clean energy technologies, and porous materials that soak up carbon.

Yet materials development today is still mostly a result of intuition and luck. The empirical process is slow and fraught with human bias and error. Artificial intelligence (AI) is poised to change that.

The combination of big data and AI is being called the "fourth industrial revolution," and its applications in materials science have soared in the past decade. Researchers have already used the AI subfield of machine learning to find new battery electrode materials and phosphors for solar cells. Identifying and developing a material for a technological application usually takes more than a decade. AI could slash that to one or two years.

Speeding up materials discovery is not the main driver for the use of AI in materials research, though. According to Gerbrand Ceder, professor of materials science and engineering at the University of California, Berkeley, "AI is best for solving problems that we don't understand and things that we don't know how to predict. Materials synthesis is a perfect example. A researcher could think of a new compound for a catalyst or electrode material, said Ceder, "but we don't have any predictive theory to tell us how to make the compound or if it can exist and be made at all. People have historical knowledge or intuition, and have made hundreds of thousands of compounds. Machine learning could find patterns in these data and learn the rules of synthesis."

Materials scientists have used computational techniques for decades. *Ab initio* computing, for instance, simulates the behavior of molecules by solving the equations that govern material properties at the atomic scale. Researchers can use it to calculate the properties of thousands of virtual compounds without stepping into the laboratory. Density functional theory (DFT) is a computational modeling method that calculates the energy of atomic structures and uses that to estimate material properties. The Materials Project aims to use *ab initio* methods to compute the properties of all known and predicted compounds.

But DFT calculations are slow and computationally intensive. In addition, DFT is limited to picosecond time scales and can only simulate about 200 atoms, said Nongnuch Artrith, a research scientist at Columbia University.

Machine learning can overcome these limitations. Machine learning comprises statistical algorithms and models that can correlate data in large datasets to decipher patterns and underlying rules, which it can then use to make predictions. These computer models learn from existing data—large, cleanly labeled data sets are a must—and their performance improves with training. In image recognition, for instance, computer scientists use millions of labeled pictures to train neural networks to recognize objects. This is how social media sites recognize faces and autonomous vehicles navigate.

"The idea is to train models on DFT reference calculations of small structures, and then use these models to run molecular dynamics simulations," said Artrith. Using this method, she and her colleagues study structures of thousands or millions of atoms at nanosecond scales. This allows them to simulate battery materials containing a large number of atomic species, such as lithium, oxygen, different transition metals, and carbon.

Researchers are also using machine learning to predict material properties by skipping *ab initio* calculations entirely. Instead, they create mathematical models to represent molecules based on size and number of atoms, and their relative locations and distance, said Gus Hart, physicist at Brigham Young University. Then the model is trained on data sets to map materials to their properties.

In 2018, Shyue Ping Ong, a nanoengineer at the University of California, San Diego, and colleagues showed that neural networks could use just two parameters—Pauling electronegativity and ion radii—to predict the formation energies of garnet and perovskite crystals. The model could accurately screen thousands of garnet and perovskite candidates and identify stable compositions. The researchers have also developed graph-based models

Shyue Ping Ong, University of California, San Diego, USA
Prachi Patel, prachi@lekh.org

trained on data from tens of thousands of crystals in the Materials Project database to predict formation energies, bandgaps, and elastic moduli of crystals more accurately than DFT calculations performed on a much larger data set.

Credit: Shyue Ping Ong, University of California, San Diego.

Another exciting use of machine learning is to create new materials from scratch. The idea is to give a trained model a set of target properties for an application, and let it scan millions of possible atomic configurations for compositions that are likely to meet the target. Using this inverse design principle, Jakoah Brgoch and colleagues at the University of Houston developed an algorithm that could predict the properties of almost 120,000 inorganic compounds to shortlist two dozen promising phosphors for white LEDs. When they synthesized and tested one of the compounds predicted computationally, sodium barium borate, they found that it had a quantum yield of 95 percent and was extremely thermally stable.

Developing new materials is a necessity for companies that rely on advanced materials to produce consumer products. "They have a need to innovate materials faster," said Bryce Meredig, chief scientific officer and co-founder of Citrine Informatics, one of several emerging startups combining AI and big data to accelerate the pace of making new materials.

"AI can be powerful, but it relies on having a large amount of data," he said. "Over the past decades, companies have accumulated data and expertise on material sets, but not in a highly structured and organized form that is accessible to AI."

Citrine helps its customers bring all their materials-related data into a single software system; develops AI software that can learn from this data; and works with the company to consistently integrate new data generated in labs and simulations into the platform.

Another startup, MaterialsZone Ltd. in Israel, specializes at correlating processing parameters, structures, and functionality. The Bar-Ilan University spin-off uses this and similar techniques in discovering new photovoltaic (PV) absorbers for future devices. Data-mining tools analyze large PV material data sets generated by high-throughput methods. The deposition process parameters then model PV performance, providing a path to higher performing solar cells.

"The properties calculated on the microscale for many energy-related materials may present differently on the macroscale," said Assaf Anderson, CEO of MaterialsZone. "In addition, lab-fabricated materials are never as ideal as the results of theoretical calculations, and the big challenges are to include and forecast effects of scaling-up, reproducibility, lifetime, and cost reductions."

AI can affect materials research in many other smaller ways too, said Meredig. It can help curate databases or help scientists process data more efficiently, "say by solving crystal structures from x-ray diffraction patterns, which can take hours and days and may not be solvable by a person."

Yet, for all the excitement about machine learning in materials science, "nobody has shown the 'killer app' yet," said Ceder. That invaluable use for AI could stem from its ability to read and process text, he said. "Can you feed a million research papers to a machine, and can it learn science from that? Evidence shows that at a primitive level, it can, just like auto-predicting the next word in a text message."

Text mining has produced some incredible results. Ceder and his colleagues, for instance, trained a neural network on three million abstracts of research papers about a compound used for solar cells published before 2005. The model predicted that the material would be used in thermoelectric applications: turns out both solar cells and thermoelectric applications require similar band gaps. Indeed, research papers on the thermoelectric applications of the material were published after 2005.

In another example, researchers at Haverford College had undergraduate researchers digitize their lab notebooks, recording reaction conditions that worked and did not work for the crystallization of vanadium selenites. The researchers were able to use the data to train a model that could predict the success of a new set of reaction inputs with 89% accuracy.

Some researchers, such as Elsa Olivetti from the Massachusetts Institute of Technology, are now tackling a more complicated task: using AI to extract materials recipes from research papers in order to predict synthesis of new compounds. Ceder believes that using text mining to predict material synthesis could be a game-changer. "Could we then create materials with properties that we don't know how to predict yet, like superconductivity?" he said.

The use of AI in materials science is still nascent, held back by a lack of data. "Materials data tend to be small and sparse," said Meredig. "You have lots of small data sets that are often siloed." In addition to more data, there is also a need for metadata—a summarized basic description of data that makes it easier to use. "Data is only as good as its metadata," said Ceder.

To see a radical change in the material development process, AI should be tightly integrated into the workforce, said Meredig. "We need to make AI a copilot in materials research." And that means training and educating the next generation of materials scientists and engineers to understand and utilize these tools. "Experimental material scientists need to know scanning electron microscopy but also AI tools." □