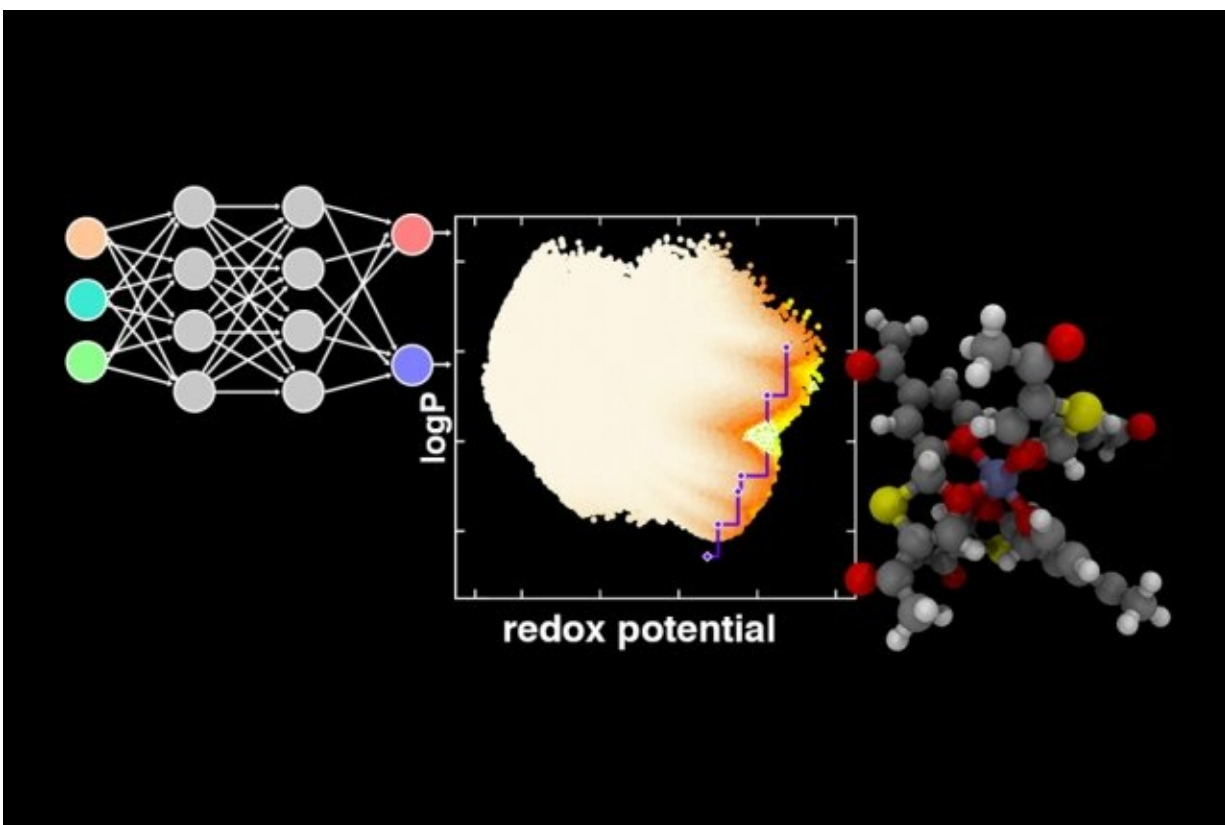# Neural networks facilitate optimization in the search for new materials

March 26 2020, by David L. Chandler

An iterative, multi-step process for training a neural network, as depicted at top left, leads to an assessment of the tradeoffs between two competing qualities, as depicted in graph at center. The blue line represents a so-called Pareto front, defining the cases beyond which the materials selection cannot be further improved. This makes it possible to identify specific categories of promising new materials, such as the one depicted by the molecular diagram at right. Credit: Massachusetts Institute of Technology

When searching through theoretical lists of possible new materials for particular applications, such as batteries or other energy-related devices, there are often millions of potential materials that could be considered, and multiple criteria that need to be met and optimized at once. Now, researchers at MIT have found a way to dramatically streamline the discovery process, using a machine learning system.

As a demonstration, the team arrived at a set of the eight most promising materials, out of nearly 3 million candidates, for an energy storage system called a [flow battery](#). This culling process would have taken 50 years by conventional analytical methods, they say, but they accomplished it in five weeks.

The findings are reported in the journal *ACS Central Science*, in a paper by MIT professor of chemical engineering Heather Kulik, Jon Paul Janet Ph.D. '19, Sahasrajit Ramesh, and graduate student Chenru Duan.

The study looked at a set of materials called transition metal complexes. These can exist in a vast number of different forms, and Kulik says they "are really fascinating, functional materials that are unlike a lot of other material phases. The only way to understand why they work the way they do is to study them using quantum mechanics."

To predict the properties of any one of millions of these materials would require either time-consuming and resource-intensive spectroscopy and other lab work, or time-consuming, highly complex physics-based computer modeling for each possible candidate material or combination of materials. Each such study could consume hours to days of work.

Instead, Kulik and her team took a small number of different possible materials and used them to teach an advanced machine-learning neural network about the relationship between the materials' chemical compositions and their physical properties. That knowledge was then

applied to generate suggestions for the next generation of possible materials to be used for the next round of training of the neural network. Through four successive iterations of this process, the neural network improved significantly each time, until reaching a point where it was clear that further iterations would not yield any further improvements.

This iterative optimization system greatly streamlined the process of arriving at potential solutions that satisfied the two conflicting criteria being sought. This kind of process of finding the best solutions in situations, where improving one factor tends to worsen the other, is known as a Pareto front, representing a graph of the points such that any further improvement of one factor would make the other worse. In other words, the graph represents the best possible compromise points, depending on the relative importance assigned to each factor.

Training typical neural networks requires very [large data sets](#), ranging from thousands to millions of examples, but Kulik and her team were able to use this iterative process, based on the Pareto front model, to streamline the process and provide reliable results using only the few hundred samples.

In the case of screening for the flow battery materials, the desired characteristics were in conflict, as is often the case: The optimum material would have high solubility and a high energy density (the ability to store energy for a given weight). But increasing solubility tends to decrease the energy density, and vice versa.

Not only was the neural network able to rapidly come up with promising candidates, it also was able to assign levels of confidence to its different predictions through each iteration, which helped to allow the refinement of the sample selection at each step. "We developed a better than best-in-class uncertainty quantification technique for really knowing when these models were going to fail," Kulik says.

The challenge they chose for the proof-of-concept trial was materials for use in redox flow batteries, a type of battery that holds promise for large, grid-scale batteries that could play a significant role in enabling clean, renewable energy. Transition metal complexes are the preferred category of materials for such batteries, Kulik says, but there are too many possibilities to evaluate by conventional means. They started out with a list of 3 million such complexes before ultimately whittling that down to the eight good candidates, along with a set of design rules that should enable experimentalists to explore the potential of these candidates and their variations.

"Through that process, the neural net both gets increasingly smarter about the [design] space, but also increasingly pessimistic that anything beyond what we've already characterized can further improve on what we already know," she says.

Apart from the specific transition metal complexes suggested for further investigation using this system, she says, the method itself could have much broader applications. "We do view it as the framework that can be applied to any materials design challenge where you're really trying to address multiple objectives at once. You know, all of the most interesting materials design challenges are ones where you have one thing you're trying to improve, but improving that worsens another. And for us, the redox flow battery redox couple was just a good demonstration of where we think we can go with this machine learning and accelerated materials discovery."

For example, optimizing catalysts for various chemical and industrial processes is another kind of such complex materials search, Kulik says. Presently used catalysts often involve rare and expensive elements, so finding similarly effective compounds based on abundant and inexpensive materials could be a significant advantage.

"This paper represents, I believe, the first application of multidimensional directed improvement in the chemical sciences," she says. But the long-term significance of the work is in the methodology itself, because of things that might not be possible at all otherwise. "You start to realize that even with parallel computations, these are cases where we wouldn't have come up with a design principle in any other way. And these leads that are coming out of our work, these are not necessarily at all ideas that were already known from the literature or that an expert would have been able to point you to."

"This is a beautiful combination of concepts in statistics, applied math, and physical science that is going to be extremely useful in engineering applications," says George Schatz, a professor of chemistry and of chemical and biological engineering at Northwestern University, who was not associated with this work. He says this research addresses "how to do machine learning when there are multiple objectives. Kulik's approach uses leading edge methods to train an artificial neural network that is used to predict which combination of transition metal ions and organic ligands will be best for redox flow battery electrolytes."

Schatz says, "This method can be used in many different contexts, so it has the potential to transform machine learning, which is a major activity around the world."

This story is republished courtesy of MIT News (web.mit.edu/newsoffice/), a popular site that covers news about MIT research, innovation and teaching.