

Machine learning techniques advance experimental design

November 1 2022, by David Silverman



Credit: Pixabay/CC0 Public Domain

Imperial and chemical company BASF will reveal new techniques for optimizing experimental design at the machine learning conference NeurIPS.

Three papers outlining new machine learning techniques that address important needs in the [chemical industry](#) have been judged groundbreaking enough to win acceptance at the NeurIPS conference, one of the most competitive international venues for research in machine learning.

The techniques, developed as part a wide-ranging partnership between Imperial and BASF, are designed to help research and development (R&D) scientists in chemistry and other fields improve [industrial processes](#) with minimal trial and error by predicting which experiments will return the most useful results. They could also help automate the R&D process. These advances are expected to help accelerate the development of innovative new chemical products and more efficient and sustainable methods of manufacturing.

Professor Ruth Misener, BASF/RAEng Research Chair in Data-Driven Optimization in Imperial's Department of Computing, said, "Our research with BASF is helping digitally transform the chemical industries by adapting cutting-edge experimental design techniques to sector-specific requirements. This will help create better and more sustainable chemistry. The research is also making important theoretical advances in machine learning. The acceptance into NeurIPS is a marker of the recognition our work is receiving from the wider machine learning community."

Dr. Christian Holtze, Academic Partnership Developer and Senior Research Engineer at BASF, said, "At BASF we see digitalization as key for strengthening our role as a leader in R&D in the chemical industry, and addressing the industry's pressing needs for sustainability and resilience. We have long valued Imperial's cutting-edge expertise in this area. An outstanding example is our work with Professor Ruth Misener and colleagues, who share our vision of developing disruptive digital capabilities for tomorrow's chemical industry."

Optimizing R&D

Industrial chemists routinely conduct experiments to develop high performing products such as coatings and batteries, maximize the purity of the chemicals, and minimize the material and [energy costs](#) of producing them. This requires testing numerous combinations of ingredients and reactor settings such as flow rates and temperatures.

Since the process of experimentation is itself costly, chemical companies aim to design these experiments to do the best possible job of optimizing a manufacturing process in a finite number of experimental iterations.

With this in mind, it does not make sense to decide in advance which parameter values to test in each iteration of an experiment. Instead, chemists typically carry out a handful of initial iterations with differing values and use the results to form a rough prediction about which settings—for example, which temperatures—will provide the best performance. Further iterations are then designed on-the-fly to progressively improve on the precision and accuracy of the first prediction.

A more sophisticated version of this iterative process relies not on human intuition but on optimization algorithms. In Bayesian optimization, an algorithm combines [experimental data](#) with a statistical background assumption known as a Gaussian process to estimate the mathematical function that links the experimental variables to the manufacturing performance. This estimate, which starts out highly approximate, is not represented as a certainty but as a [probability distribution](#) over a range of possible functions.

The algorithm's aim is to find the settings that yield the best manufacturing performance, and the probability distribution helps it do this by enabling the algorithm to predict which experimental settings

have the greatest chance of yielding better performances. Experimental scientists in a range of academic and industrial fields are increasingly using machine learning algorithms that employ this method.

"The algorithms are often superior to human intuition as it's really hard for people to see what's going on when there are many variables," says Dr. Robert M Lee, a machine learning specialist at BASF who carried out the research with colleagues at Imperial and BASF colleagues Dr. Behrang Shafei and Dr. David Walz.

"The other advantage they offer is automation. In most cases we still have a human in the loop, but we have a couple of cases where we can close that loop, meaning you can hit go, focus your attention on something else, and come back to some good results."

Chemistry advances

While optimization of experimental design is already a successful field, industrial chemistry is still grappling with practical constraints that are hard for optimization algorithms to account for. These include:

- The cost of changing physical variables—for example temperature, which is easier to change a little than a lot.
- The fact that some types of experimental data come back faster than others and chemists need to make [experimental design](#) decisions before receiving all of them, known as asynchronous batch.
- Multiple objectives, for example the need to minimize the cost of manufacturing chemicals while also maximizing quality and sustainability.
- The combination of continuous variables like temperature with categorical variables like on and off.

- Multi-fidelity, or the fact that some data sources are more trustworthy than others.
- Input constraints, such as facts known independent of experimentation, for example that chemical components expressed as percentages must add up to 100%.

The new techniques developed by researchers from Imperial and BASF accommodate many of these constraints. One paper provides an optimization technique that compromises between multiple objectives, while another accounts for input constraints and varied types of variables. A third accounts for the costs of changing variables, multi-fidelity and asynchronous batch.

The papers were authored by Ph.D. students and academics in Imperial's Departments of Computing and Mathematics with R&D scientists at BASF. The research was funded by BASF and the Engineering and Physical Sciences Research Council via the Center for Doctoral Training in Statistics and Machine Learning.

New frontiers in machine learning

While motivated by practical needs in chemistry, the research has translated these into mathematical techniques with general applications in machine learning and the potential to make a large impact outside the chemical industries themselves.

Mr. Alexander Thebelt, a Ph.D. student in Computing and lead author of one paper said, "The machine learning community is getting more and more interested in these kinds of problems. A lot of the applications that are notoriously successful in machine learning are cases where you have loads of data, and you can just throw as much data as you want at the algorithms to make them work. This is not the case in all fields. If we can use our techniques to discover new materials that could have a huge

impact on humanity and industry. It could be as big as any of the successes we've seen in machine learning so far."

Dr. Lee said it is not unusual for the application of problems from industrial chemistry to open up new theoretical approaches in machine learning research. "People from [machine learning](#) traditionally think of data as either in a table, or maybe a picture or text. As soon as you come and say, my inputs are not these things, they're actually molecules, then people say, oh, interesting, I can represent a molecule as a graph with nodes and edges and do all sorts of clever stuff."

"Our partnership with Ruth [Misener] is extremely valuable because her group has one foot in chemical engineering and one in AI, which is essential in this kind of work. It's refreshing for us to collaborate with academia; it gives you a different way of looking at things. It's stuff we couldn't develop in-house."

More information: Jose Pablo Folch et al, SnAKE: Bayesian Optimization with Pathwise Exploration, *arXiv* (2022). [DOI: 10.48550/arxiv.2202.00060](https://doi.org/10.48550/arxiv.2202.00060)

Alexander Thebelt et al, Tree ensemble kernels for Bayesian optimization with known constraints over mixed-feature spaces, *arXiv* (2022). [DOI: 10.48550/arxiv.2207.00879](https://doi.org/10.48550/arxiv.2207.00879)

Ben Tu et al, Joint Entropy Search for Multi-objective Bayesian Optimization, *arXiv* (2022). [DOI: 10.48550/arxiv.2210.02905](https://doi.org/10.48550/arxiv.2210.02905)

Provided by Imperial College London

Citation: Machine learning techniques advance experimental design (2022, November 1)

retrieved 25 March 2023 from <https://techxplore.com/news/2022-11-machine-techniques-advance-experimental.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.