Quantitative Machine Learning for the Chemical Sciences

William H. Green, MIT Dept. of Chemical Engineering

Machine Learning to predict text or to recommend content to online users, based on huge training datasets, is increasingly important in commerce. In this talk, we discuss how modern Machine Learning techniques can be adapted to help solve chemical problems. Machine Learning can be used to predict discrete quantities (e.g. "what catalyst to use to synthesize a new molecule?") or to predict numbers, with significant digits and error bars, from chemical structures (e.g. "how soluble would this molecule be in methanol at 40 C?" or "what is the barrier height of this proposed reaction"?). Chemical data is much more difficult to acquire than the types of data used to train most machine learning models, so a large amount of effort needs to be devoted to collecting and curating the data. Chemical space is huge, but chemical datasets are often relatively small, so methods for augmenting the limited experimental data with physical models are important. Usually we are extrapolating outside of the limited training data to predict new molecules or reactions, so conventional approaches that assume interpolative fits give overly optimistic estimates of the uncertainty of predictions. Despite all these challenges, we have made significant progress, building many useful models. Indeed, in some cases the computer predictions are more accurate than typical experimental measurements, and without the need to spend time synthesizing, purifying, and measuring the new molecules. Recently, we have used these computer models to drive "active learning" cycles to invent new molecules with improved performance for certain applications, and to also design and execute robotic synthesis and measurements of the new molecules. The current status of this rapidly growing field is reviewed, and some unresolved challenges are discussed.

About the speaker: William H. Green is the Hoyt C. Hottel Professor at MIT. He earned his B.A. at Swarthmore College and his Ph.D. at the University of California. After postdocs at Cambridge University and the University of Pennsylvania, he worked at Exxon Research & Engineering for 6 years, then joined the faculty at MIT in 1997. He has published more than 300 journal articles cited more than 20,000 times. He is a Fellow of the American Association for the Advancement of Science and a Fellow of the Combustion Institute, and has received several major awards including American Chemical Society's Glenn Award in Energy & Fuels and the American Institute of Chemical Engineers' R.H. Wilhelm Award in Reaction Engineering.