Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

AI3SD Interview with Dr Reinhard Maurer
18/11/2019
AI3SD Conference 2019

Michelle Pauli
Michelle Pauli Ltd

20/08/2020

AI3SD-Interview-Series:Report-5

AI3SD Interview with Dr Reinhard Maurer
AI3SD-Interview-Series:Report-5
20/08/2020
DOI: 10.5258/SOTON/P0028
Published by University of Southampton

Network: Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery
This Network+ is EPSRC Funded under Grant No: EP/S000356/1

Principal Investigator: *Professor Jeremy Frey*
Co-Investigator: *Professor Mahesan Niranjan*
Network+ Coordinator: *Dr Samantha Kanza*

Contents

Dr Reinhard Maurer, Assistant Professor in Computational Chemistry, University of Warwick, takes the AI3SD Q+A.

1 Interview Details

Title	AI3SD Interview with Dr Reinhard Maurer
Interviewer	MP: Michelle Pauli - MichellePauli Ltd
Interviewee	RM: Dr Reinhard Maurer - University of Warwick
Interview Location	AI3SD Conference 2019
Dates	18/11/2019

2 Biography

Figure 1: Dr Reinhard Maurer

Reinhards research focuses on the theory and simulation of molecular reactions on surfaces and in materials. Reinhard studies the structure, composition, and reactivity of molecules interacting with solid surfaces. Reinhards goal is to find a detailed understanding of the explicit molecular-level dynamics of molecular reactions as they appear in catalysis, photochemistry, and nanotechnology. Members of Reinhards research group develop and use electronic structure theory, quantum chemistry, molecular dynamics, and machine learning methods to achieve this.

3 Interview

MP: You have an AI3SD-funded project, *Deep-learning-enhanced quantum chemistry: pushing the limits of materials discovery*, that recently delivered its final report. What's the background to the project?

RM: The idea behind it is that in computational chemistry, basically, everything we do revolves around solving quantum mechanical equations, like the Schrödinger equation and others, in order to predict the structural and electronic properties of materials, to understand what molecules do and how they react.

That's actually quite tedious. And costly. It involves high performance computing facilities and it drains a lot of energy and resources. So we're interested in finding more efficient ways to solve these equations or, in fact, replace the need to solve them by finding machine learning models to do that.

MP: What was the aim of the project, in a nutshell?

RM: The goal was to find a way to integrate machine learning into established quantum chemical calculation methods, enabling us to combine the accuracy of high level approximations of quantum chemistry with the efficiency of much more simple approximations, so we have a method that's both accurate and fast at solving very, very large systems.

'Large' in this context means many thousands of atoms, and thousands and thousands of electrons in a system. It's a size that allows us to study very complex chemistry in certain materials. My personal interest in this, and what my research group mostly focuses on, is surface chemistry and catalysis: we're interested in looking at light and electric field-driven chemistry on electrodes and metal surfaces.

Our first aim, based on some preliminary work, was to develop a machine learning model and artificial intelligence algorithm that allows us to capture the wave function – the central quantity of quantum mechanics – and integrate this into an established software package for quantum chemistry so that the result is very accessible to users, and then apply it to some systems.

MP: What have you achieved?

RM: We haven't got as far as we wished, as is often the case with such a project, but we have set up the model to predict the wave functions, with a very big part of the effort being the generation of training data. Artificial intelligence (AI), algorithms and machine learning need to learn from data in order to predict properties.

It turns out that one of the most challenging aspects is, "How do you curate the training dataset, what goes into this training dataset and how do you ensure that you're not missing out certain things?" So we generated the training dataset with this computationally expensive data and implemented the algorithm, which is what Dr. Adam McSloy, the postdoc on this project, finished last month. And now we're in the process of training the models so that they can predict the wave functions, and integrals between wave functions, and then the next step is the integration into the code.

We hope that by the end of the year we'll have a fully integrated code that can basically go from training data all the way to wave function prediction, to integrating and using it within a software package that's accessible to a very large user base around the world.

MP: Has anything surprised you so far?

RM: It was surprisingly hard to curate such a training dataset. When you want to have a model that correctly predicts the wave functions and the properties of certain molecules, how do you ensure that you're not missing bits and pieces? That's something that took us a while to figure out, to have recipes for creating the training data set.

The other thing I found challenging is that the code base we're using is written in a machine learning framework called TensorFlow. It's quite common but it's very hard to get into. The way you have to construct the code is unusual, because everything has to be described in terms of matrix multiplications and matrix algebra. A lot of mathematical thinking goes into "How do I make a code work within this framework, that satisfies all these provisions?"

But it's been interesting, and the postdoc has, I think, learned a lot. If nothing else, at least the postdoc has really, really learned some valuable skills which are going to be important further down the line.

MP: You've talked about both AI and machine learning. We often tend to use the terms interchangeably but how would you describe the difference between them?

RM: We typically refer to what we do as machine learning because it's a very limited analytical, nonlinear model with which we try to describe certain properties. Machine learning is quite limited in its scope, while artificial intelligence, as I perceive it, is something that has a much wider scope and is less limited to certain things.

I have not seen anything in chemistry yet that I would qualify as artificial intelligence but we might be slowly getting there. Some colleagues of mine have recently published generative machine learning models which basically allow you ask the model, "Give me any molecule that satisfies a certain property," and that, in my opinion, is already much, much wider in scope.

MP: Given the limitations you've just expressed, how would you say AI and machine learning are changing how we do science?

RM: A few years ago, I was sceptical because it felt like hype, but this project and some others – and my exposure to machine learning – has really changed it for me. The way it's now picking up in science, particularly in chemistry and even more so in computational chemistry, is so transformative that I think it will eat up everything and there will be nothing left.

I'm serious. I think that in 10 years we will reach a stage, at least in computational chemistry and theoretical chemistry, where we don't even mention machine learning anymore because it will have become embedded in everything we do and using it will be commonplace.

MP: What promise does that hold for science?

RM: Our study is among several published in the last year that really show that, in principle, anything we do in computational chemistry can benefit from machine learning, from molecular dynamics simulations, where we follow the time evolution of molecules and their dynamics, to spectroscopy, to any possible quantity we could be interested in.

At the moment, we see lots of publications coming out that show the proof of principle: "Yes, we can predict these properties using machine learning models." And yes it gives us many orders of magnitude in acceleration and allows us to understand things we otherwise wouldn't be able to understand, or makes the tool set of the computational chemist more complete.

Now we're slowly moving to the stage where we incorporate this into our daily life, into the software packages that we use, so that PhD students use it day in, day out. I think this phase will be past its peak within the next five years and then, as I said, we will slowly stop talking about it.

One of the most exciting aspects of machine learning is that it provides an additional way to bridge experiment and theory. I work a lot with experimental colleagues and we have a lot of collaborators. Very often, the questions they ask us, the problems they bring are too challenging, too hard. We don't have the methods, we don't have the capabilities. I think machine learning, allowing us to scale up and deal with more complex problems, is going to help with that.

My big fear, however, is that we might be in the process of making ourselves obsolete. Hopefully, that is going to take a few more years!

MP: And what of the intellectual debt, the idea that we get the answer but we have no idea how we got there, and does that matter?

RM: It needs to matter how you got to the answer and if it was a sensible path. Any good theory needs to be falsifiable.

Freud famously said, "Man has, as it were, become a kind of prosthetic God". You have lots of tools that allow you to improve your senses, and all the experimental equipment that we have and all the theoretical methods that we have are ways to improve our senses, and machine learning is just another tool in that toolkit. I don't really see the computer killing humanity. It's just an important tool and, if it's used wisely, then in science it's going to make a big difference.