Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

AI3SD Interview with Dr James Cumby
17/03/2021
Online Interview

Michelle Pauli
Michelle Pauli Ltd

04/01/2022

Humans-of-AI3SD:Interview-16

AI3SD Interview with Dr James Cumby
Humans-of-AI3SD:Interview-16
04/01/2022
DOI: 10.5258/SOTON/AI3SD0179
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

1 Interview Details

Title	AI3SD Interview with Dr James Cumby
Interviewer	MP: Michelle Pauli - MichellePauli Ltd
Interviewee	JC: Dr James Cumby - University of Edinburgh
Interview Location	Online Interview
Dates	17/03/2021

2 Biography

Figure 1: Dr James Cumby

Dr James Cumby: 'You can get a very good result with a very simple machine learning model'

James Cumby is a Lecturer in inorganic chemistry at the University of Edinburgh.

In this Humans of AI3SD interview he discusses making oxyfluorides, getting closer to a universal descriptor, the power of simple machine learning tools, and why early career researchers should try to find the sweet spot between 'fun' and 'funded'.

3 Interview

MP: What's been your path to where you are today?

JC: I'm a lecturer in inorganic chemistry and my research spans solid state chemistry through to condensed matter physics – that strange "materials chemistry" boundary. My main aim is to make new materials that are useful for something. Ultimately, I'm an experimental chemist at heart. My background training is entirely experimental and stems from wanting to design better materials that are going to have properties such as being good magnets, or good electronic conductors, or have interesting mechanical properties such as negative thermal expansion.

I did my undergraduate degree and PhD at the University of Birmingham, with Professor Colin Greaves, looking at structures that were synthetic analogues of minerals, substituting different metals and seeing the effect on the magnetic properties. I moved to Edinburgh as a postdoc, working on a European project looking at magnetite, the oldest known magnetic material, which still holds lots of interesting problems. When you cool it down, it undergoes what's called a phase transition. The atoms move around slightly, readjust their positions, and essentially it's a charge ordering transition. You go from a situation where two iron atoms share an odd number of electrons equally to one iron atom holding on to the spare electron, and the other being without it (although reluctantly). Ultimately I'm interested in atomic structure. What influence do different things have on it? In magnetite you can substitute atoms and change it so there is slightly less than one spare electron per pair of iron atoms. And then what happens to this transition? Does it still occur? Does it go the other way or does it do something completely different?

I use synchrotron radiation a lot, as well as neutron diffraction to look at magnetic properties and explore the underlying atomic behaviour that's giving rise to these physical properties. I believe that if you know the atomic structure of a material, that should predict its physical properties and the properties come entirely from where the atoms are.

I stayed at Edinburgh, first for a teaching-only position then, six months later, on a mixed teaching and research lectureship. This research is looking at mixing oxygen and fluorine in materials as a way of trying to control these phase transitions and develop better functional materials. Many oxide materials have been studied in the past. Metal oxides are very well known and occur in most of the devices and technologies that we're used to, such as laptops. But oxyfluorides, where you're mixing oxygen and fluorine, are much less understood and explored, but they have an extra degree of freedom. You can change how much oxygen versus fluorine and use that to control the properties.

My research is focused on how we can get better at making oxyfluorides. How can we characterise them faster? There are almost a hundred elements you can mix in various combinations. You can't just go into the lab and make an arbitrary number of these because it's just too slow. So that led to an interest in machine learning – could we use existing data on crystal structures to drive the discovery of new oxyfluorides?

MP: And that's at the heart of your AI3SD project?

I was quite surprised when I started this that a lot of machine learning work has been done in chemistry, but not so much on inorganic crystal structures. The majority is focused on small organic molecules, for instance. You can't take a crystal structure and natively feed into the existing machine learning algorithms because it doesn't work. There are too many possible arrangements of atoms and there isn't a unique representation. So then you get into rounds of featurisation or creating descriptors for these crystal structures. The more I dug into it, the more I realised that, actually, there weren't really any ideal descriptors for inorganic crystal structures. They're all quite specialised and some would do very well in some problems, but completely fail at others. There's no universal descriptor. So the idea of the project was to try and get a little bit closer towards a universal descriptor.

MP: Have you achieved that?

It's closer than it was, I think. The project started to balloon beyond its initial scope the more we started to look at this problem of, if you try and predict something about the material, do you need to know which atoms are there? Or do you need to know where the atoms are in space? The literature is split between people choosing composition and people choosing structure. They either pick the structure and try to use that to describe it, or they pick the composition and use that to describe it. They both work quite well, depending on the method that you're using. We started looking at this idea of, if you know the structure, can you predict the composition? Are they inherently linked?

We're starting to see that, actually, you can't completely separate composition and structure. There are some things that you can't ever get from the structure alone that you need to predict some of the physical properties. But in terms of the broader project scope, I think we've implemented a couple of methods that do a very good job in terms of structure only. If you ignore the composition entirely, we can do better than the existing structure-only descriptors in terms of predicting things like mechanical properties of structures. And then if you add the composition in as well, you can do even better. We're at the stage now where, having split these two ideas apart completely, we now are trying to recombine them in some way that is a little bit more universal, but depending on what problem you're looking at, you can weight one more heavily than the other.

MP: What happens next?

We've got some initial results and we're going to seek further funding to take the project forward and apply it to different problems. It has been one of these things that has opened up more questions than we started with, which is the way science works best.

MP: Has anything surprised you in this work?

When I first started looking at machine learning for this problem, I was surprised that it hadn't been done already. I just assumed that, given how popular machine learning is, somebody would have solved this problem. So I thought, "Oh, I'll take the tools that are already there and it will help me predict some oxyfluorides and that will be fine." And it's turned out to be very much a rabbit hole.

The machine learning tools themselves have been a bit surprising to me. We've tried a number of different algorithms and different models to try to solve these problems and have often found that you can get a very good result with a very simple model. So we've gone from starting off with something like linear ridge regression or non-linear ridge regression models, which do quite well, but are not perfect, and then as we started going a bit further, they became non-applicable. We took a step back and started looking at nearest neighbours and trying to predict properties based around nearest neighbours. I was absolutely astonished that what, from a machine learning perspective, is considered the most basic of basic models was actually predicting just as well as the linear regression. You don't need to go straight into neural networks and big models with thousands of parameters. You can start with something very simple.

MP: How do you think machine learning is going to change how we do science?

It all comes down to the data. If the data is there, machine learning will have an impact. If the data is not, it won't. So in a sense, I'm very lucky that crystallography and crystallographers were ahead of the curve in that they were archiving data in the 70s when big databases started.

If the data is there, it is a case of finding the tools to process it. That's where machine learning is moving in the right direction and the tools are becoming more and more powerful. For standard experimental chemists, if their primary research concerns do not involve data, then machine learning is not going to really feature on their radar.

MP: What impact has Covid-19 had on your work?

It's made me very thankful that I have a computational aspect of my research, first and foremost. From a personal point of view, it made me think a lot more about how I'm communicating with a research group. If things had just continued as normal, I probably would've had an occasional group meeting in person and sent lots of emails because that was the way things were always done. But suddenly going online made me look at the options like, can we use Teams to pull information together? Can we collaborate more online? And all these things have been beneficial, in that people are now much more connected in a way while at the same time being more isolated.

MP: What advice would you give to an early career researcher based on your experience?

When I was applying for lectureships, something I think was really valuable was spending quite a lot of time thinking about what I wanted to do. It's very easy to get caught up in the research that you're doing and the problems that you're solving. And it was actually quite nice to step back and think, ”Do I want to do this for the rest of my life? Do I want to change my research focus?” And for me, it was always, I wanted to do research. It was just what research did I want to do? And so this was why I went from fundamental condensed matter physics to slightly more to applied science. I realised, personally, it wouldn't be that fulfilling to keep looking at fundamental physics problems. I wanted to start trying to find some materials that might be useful. There's a Venn diagram of the fun things that interest you and the things that people will fund you to do, and you have to find your little niche as a researcher. I'm still not sure I've found it but perhaps you never quite find it. You sort of fumble around a bit until you get something that satisfies both of those requirements at the same time.