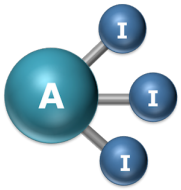

Artificial Intelligence and Augmented Intelligence for Automated Investigations for Scientific Discovery

AI3SD Interview with Professor Tim Albrecht
18/11/2019
AI3SD Conference 2019

Michelle Pauli
Michelle Pauli Ltd

06/10/2020

AI3SD-Interview-Series:Report-7

AI3SD Interview with Professor Tim Albrecht
AI3SD-Interview-Series:Report-7
06/10/2020
DOI: 10.5258/SOTON/P0030
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

Tim Albrecht, professor of physical chemistry, University of Birmingham, takes the AI3SD Q+A.

1 Interview Details

Title	AI3SD Interview with Professor Tim Albrecht
Interviewer	MP: Michelle Pauli - MichellePauli Ltd
Interviewee	TA: Professor Tim Albrecht - University of Birmingham
Interview Location	AI3SD Conference 2019
Dates	18/11/2019

2 Biography

Figure 1: Professor Tim Albrecht

Tim studied Chemistry at the University of Essen in Germany from 1995-2000. Following brief research visits at the European Joint Research Centre in Ispra in Italy and the University of California at Berkeley, Tim graduated with a Diploma in Chemistry (equivalent to a Masters degree) in early 2000. After graduating, Tim joined Peter Hildebrandt's group at the Max-Planck Institute for Radiation Chemistry (now Bioinorganic Chemistry) in 2000. Tim worked on charge transfer processes in natural and artificial heme proteins on metal surfaces using SER(R)S, single-crystal electrochemistry and electrochemical STM (in Jens Ulstrup's group at the Technical Institute of Denmark (DTU). He obtained his PhD from the Technical University (TU) of Berlin in 2003 and afterwards returned to Ulstrup's group as a postdoctoral fellow. In 2006, he moved to London to take up a lecturer position in Interfacial and Analytical Sciences in the Chemistry Department at Imperial College, where he was made Senior Lecturer in 2011 and then Reader in 2014. In 2017, Tim joined the faculty in the School of Chemistry at Birmingham University as Chair of Physical Chemistry.

3 Interview

MP: Your AI3SD-funded project is titled *'Next-next' generation quantum DNA sequencing with chemical surface design and capsule nets.* In a nutshell, what is it?

TA: Essentially, we're aiming to use capsule nets to improve tunnelling-based DNA nucleotide detection for new sequencing applications.

MP: Capsule nets?

TA: Capsule networks. CNs. A deep learning methodology. They are similar to convolutional networks, CNNs, that are hugely successful in areas such as image and speech recognition. We have already used CNNs in sensing applications, but we are only at the beginning of exploring such methods. In brief, compared to CNNs, CapsNets are said to be better at recognising image features, as in the so-called "Picasso Problem". We anticipate that CapsNets could also offer benefits in event detection and identification, even though they are usually computationally quite expensive.

MP: And what would be the resulting benefit of this new approach?

TA: If it works, it could take us another step towards much faster, and therefore cheaper approach to DNA or protein sequencing. The new DNA sequencing technology would enable a lot of new applications, for example in disease diagnostics or disease monitoring, and offer an ability to sequence not only DNA but other biopolymers, such as peptides, RNA and more.

The end-goal is to take quantum tunnelling-based biosensing and sequencing to a level that would make it the most promising contender for the label-free, 'next-next' generation sequencing of biopolymers.

MP: How far along the road are you?

TA: At the moment, we're optimising single-nucleotide detection using tunnelling. Once we've optimised the process to sufficient accuracy, we'll use it on a high-throughput platform and, all going well, that should constitute a sequencing technology.

We'll know over the coming months about the results for the single nucleotide optimisation. Beyond that, to a point of implementing the new technology, we're probably looking at a five-year time frame.

MP: What part does the AI3SD funding play?

TA: The new data analysis method is what forms the basis of the AI3SD project. The funding has been a stepping stone, in a sense, in that it accelerated the artificial intelligence- (AI) based approaches. We've been able to attract further support for a PhD student, for example, who is now working on the acquisition of experimental data. It's a challenge to fit the time frame of the AI3SD project with the diversity of incorporating both AI development and acquiring experimental data.

MP: Would this have happened without the AI3SD funding?

TA: Probably not at this point in time. This is an ongoing, longer-term activity but it was helped by the fact that not only was the funding available at this point in time, I also had the right person in place who is very interested in AI-based techniques and brought a lot of basic knowledge with him. Basically it meant we hit the ground running.

MP: Have any of your findings so far surprised you?

TA: What's certainly reassuring is the robustness of the data analysis methodologies we've used: they have consistently shown high performance, albeit with simulated data, with accuracies in the region of 99%. That is what you might call our internal benchmark.

But so far this is with simulated data. The test is going to be when we use real data and, as I said, we're working on acquiring sufficient experimental data at the moment.

MP: How do you feel AI has changed scientific discovery?

TA: A lot. Compared to what it could achieve, I think it hasn't changed the landscape that much yet but it has thrown up some very fundamental questions as to how we do science and what is actually required. And my feeling is that there's a lot more potential because the AI-based techniques are used at the moment in particular contexts and so far I don't think there's any sort of overarching integration into the scientific endeavour.

For example, in our case it helps to improve a particular sensing problem. It would help other people to find a particular type of proof material. But it's not routine. I feel we're not using all the knowledge that's out there and taking advantage of it in our daily work. To some extent that has to do with the heterogeneity of the data that's available: the difficulty in assessing what is good quality, what's bad quality. It's probably also due to a lack of standardisation.

Of course, AI also challenges the fundamental notion of scientific work. The classical approach would be that we have a hypothesis, we test that hypothesis and then modify it and work towards a particular goal. By contrast, some of the AI-based approaches are simply brute force. For example, in the commercial world, when you've got automated synthesis, where people just run through a large number of different compositions and then test the result, you don't necessarily care why your particular proposition worked, as long as it works well. The 'why' is significantly less important than the fact that it actually does work.

Or in medicine: as long as you save the patient, you don't necessarily wonder how it worked and why, and I think you're not going to ponder for a long time over that question.

On the other hand, if you're in court and, based on an analytical result, someone is found guilty or not guilty, it would be uncomfortable, at least, not to know how an algorithm arrived at a particular answer.

I think DNA sequencing is a very good example of where deep learning-based approaches have been used since it started, in terms of RNNs, for example. But then you've got a particular letter code and you basically use that to improve prediction.

But I think it's very different when it comes to developing new things, new ideas. What's a

new idea? It's a surprising idea. We find something creative because we haven't thought about it before or we've put something into a new context. Maybe transfer learning-based approaches can do that now but whether humans might be surprised by that and find it creative. . . ? However, the mechanism by which you arrive at that goal might be very different than what a human being would do. That's why it challenges the way we do things.