Nicos Starreveld
Platform Wiskunde Nederland
Korteweg–de Vries Instituut voor Wiskunde
Universiteit van Amsterdam
n.j.starreveld@uva.nl

Interview Mateja Jamnik

Bringing mathematical intuition into AI models

Professor Mateja Jamnik is a full Professor of Artificial Intelligence in the Department of Computer Science and Technology (Computer Laboratory) at the University of Cambridge, UK. Next to her research she serves on the UK Home Office Science Advisory Council and had served previously as Specialist Adviser to the House of Lords Select Committee on Artificial Intelligence. She was one of the organizers of the workshop Mechanization and Mathematical Research at the Lorenz center last September. During a visit in Cambridge, I grasped this opportunity and interviewed her about her career, research, and view on the future.

Thank you very much for making time for this interview. It is very nice to be here in Cambridge. It is clear from your work that you became interested in both AI and mathematics at an early stage. Could you maybe tell us how these interests have grown?
"I was studying mathematics for my undergraduate degree in Canada, where I also took some computer science courses. In Canada you are free to combine various subjects. I was always interested in intuitive proofs of mathematical results, and at some point, I thought "Hey, how can we model this wonderful intuition that mathematicians have, or this really accessible beautiful and elegant mathematical reasoning, on machines?" I decided then to pursue a postgraduate qualification into computer science in Cambridge, and in 1995, I applied for a PhD position in Edinburgh in the group of Alan Bundy. At that time, the University of Edinburgh was the only university worldwide with a department focusing on AI. This is interesting because that period was considered the AI winter. AI was deeply unpopular because of all the deep promises made earlier that were not

Mateja Jamnik

delivered. At that time, funding for such research was reduced and it was not cool to work on AI."

Could you describe globally what your PhD research was?

"During my PhD I developed computational models that capture human mathematical thinking. I was trying to show whether it is possible to capture intuitive mathematical proofs in a computer. I was primarily looking at schematic diagrammatic proofs. I built a theorem prover, called DIAMOND, which could generalize visual proofs; at the same time it would provide correctness guarantees. The user should provide the main idea of the proof by presenting the proofs of some specific instances. Afterwards, the solver would generalize the proof for the general case. I showed that you can devise completely formal proofs with visual, diagrammatic inference steps. This dispelled the long-held assumption that diagrams cannot be formal."

Before continuing with the interview, let us take a moment to dive into the details behind such an automatic theorem prover. Suppose you want to prove the identity $1+3+\ldots+(2n-1)=n^2$.

A mathematician could deploy various methods. For example, you could prove this result by using induction, or by using the standard idea of Gauss which we learn at school. But there is also a visual diagrammatic proof. If you arrange dots in an $n \times n$ grid and count in two different ways, one by taking the size of the grid and one by summing over the n L-shapes. Then you obtain the same result. This second proof using a grid and L-shapes can be automated:

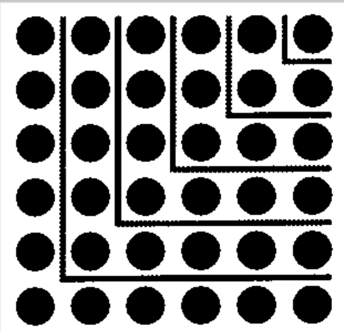

Figure 1 A grid of 6^2 dots divided in 6 L-shapes each one consisting of ($2i$-1) dots, i=1,2,…,6.

The main idea is to work with suitable objects generated from a class `diagram`, similar to an object-oriented programming language. The main building block is `dots`. Such objects can be `squares` (like in the picture above), `ells`, `rows`, `columns`. For example, `diagram(square, 6)` represents a square of size 6×6 and `diagram(ell, n)` an L-shape of $2n-1$ dots. Various operations can be applied to these objects. This enables a computer to manipulate them and find relations behind them. For example, applying the operation `L_cut` once on a `diagram(square, n+1)` will yield two smaller diagrams, namely `diagram(ell, n+1)` and `diagram(square, n)`. Such steps can be carried out by a program. This object-based approach makes the formalization of such a proof possible.

The most important concept behind this method is the constructive ω-rule, which requires there to be an effective procedure, like a recursive program. The ω-rule is infinite and here we rely on the constructive part, which makes it programmable. This works as follows. Firstly, the user needs to prove some special cases of the problem, say for $n=2,3$ and 10. From these special cases we extract a set of rules R which are applied on the problem for a general n, we call this set of rules `proof(n)`. The last step is to let the machine verify, following formal logic rules, that `proof(n)` proves the statement by using meta-induction on n. In practice, such a proof entails that the set of rules in `proof(n)` can establish that an equality between two lists of diagrams is true. In the example above, the computer would prove that a square of size n and a collection of n L-shapes each having size $2i-1$ are equal. Equality here is defined in a specific way as equality between diagrams using formal logic. We recommend the interested reader to have a look at [1] for the exact details and [2] for some ideas on the nature of a mathematical proof. Both are very interesting and accessible articles.

Since the time of your PhD, which was during an AI winter as you mentioned, many things have changed and keep changing daily.

That is definitely true! In this field, the tempo at which things change is very high.

How has this affected your research? What have you been working on in the past years?

As mentioned before, when I started my PhD, we were in an AI winter. But then things changed because technology caught up, and concepts that were discovered earlier started being implemented. The focus shifted to neural networks and statistical probabilistic AI, which is known today as machine learning. Nowadays, large language models (LLMs) are extremely powerful, and they can do a lot of informal mathematics. They can also come up with informal mathematical arguments and solutions. But they can't provide correctness guarantees, and people realize that you can't just deploy such methods in safety-critical situations. Especially when decisions about people are involved, like in medicine. In the past years, we worked hard to combine the best methods from the symbolic world, which are rule-based systems and can provide correctness guarantees, and the best from the statistical world, where learning can occur. Methods combining symbolic and statistical approaches are called neurosymbolic. That is exactly where my work is.

Could you give us some concrete insights into how such neurosymbolic methods work?

You should see a neurosymbolic solver as an ecosystem of a theorem prover, an automatic proof checker, and a conjecturer. You start with a preselected target set of problems and theorems that the neurosymbolic solver can't prove yet. By training the solver, you aim to increase the number of target theorems it can prove. You train the solver by using a dataset of theorems and their proofs. The idea is that the conjectures at the beginning are simple, and they are not intended directly to prove target theorems, but to train the solver to gradually get better at solving target theorems. The conjectures that improve the solver the most with respect to target theorems are used to generate new similar but slightly harder conjectures. In this way, the conjectures progressively get harder and closer to the target theorem, and the solver gets progressively better and better at proving target theorems.

The bottleneck currently is that we don't have enough data, in the form of theorems and their proofs, which can be used to train the solver. In such a dataset both the theorems and their proofs must be written in a formal language, so that the machine can process them. A future step is to also put the human mathematician in the loop. A human can decide and evaluate which conjectures are interesting for the neurosymbolic conjecturer. We envision this ecosystem where the human mathematicians would be able to use it to explore mathematics, that is where discoveries and creativity come into play.

How is the AI landscape at the moment?
I have been working for the last four to five years on neurosymbolic solvers. Recently, we pioneered a technique called 'autoformalization'. This technique uses LLMs to formalize informal mathematics, so that it can be processed by theorem provers like Isabelle and Lean. Isabelle is the largest and most comprehensive interactive theorem prover out there. The landscape is changing daily which creates some pressure. We need to stay pragmatic, stay enthusiastic and believe in the quality of our work. We should also know where our strengths lie. For example, it makes no sense for us to engage in large computations and large datasets because we can't compete with tech companies like OpenAI. We must stay creative and keep investing in our strengths. Neurosymbolic solvers, for example, were developed by academics. Afterwards companies have picked up this concept and managed to scale them up. Academics have the know-how, the freedom to look into fundamental problems, and the space to be creative. Let us also not forget that if industries pick something from research and scale it up, that is a sign of success.

Besides research you have been actively involved in advisory boards helping the government. How did you experience being part of such committees?
I participated in the past in the Select Committee on Artificial Intelligence [3]. The goal was to advise the House of Lords about policy regarding AI and how it can impact society and the economy. The government had commissioned the House of Lords to write a report regarding the impact of AI on society and to come up with recommendations for the government. Currently, I am part of the science advisory board of the Home Office, responsible for AI. This provides recommendations to the ministers and the government on a permanent basis. These are very interesting efforts where you can make a real impact.

What do you think are the future perspectives of Europe in the AI landscape?
In Europe, we have a lot of expertise and know–how. Also, we are highly connected, collaborations are easy, and we have funding for joint projects. We should capitalize on that. Europe is also very strong in fundamental research, and we can lead where AI is going in terms of regulation and safety. I think that these are very important points where Europe is strong; we should capitalize on them and keep investing in creativity and research.

References

1. On Automating Diagrammatic Proofs of Arithmetic Arguments, M. Jamnik, A. Bundy and I. Green, *Journal of Logic, Language and Information* 8: 297–321, 1999.
2. A. Bundy, M. Jamnik, & Fugard, A. 'What is a proof?', *Philosophical Transactions A: Mathematical, Physical and Engineering Sciences*, vol. 363, no. 1835, pp. 2377-2391, 2005.
3. AI in the UK: ready, willing and able?, *HOUSE OF LORDS Select Committee on Artificial Intelligence, report of Session 2017–19*.

Illustration: Ryu Tajiri