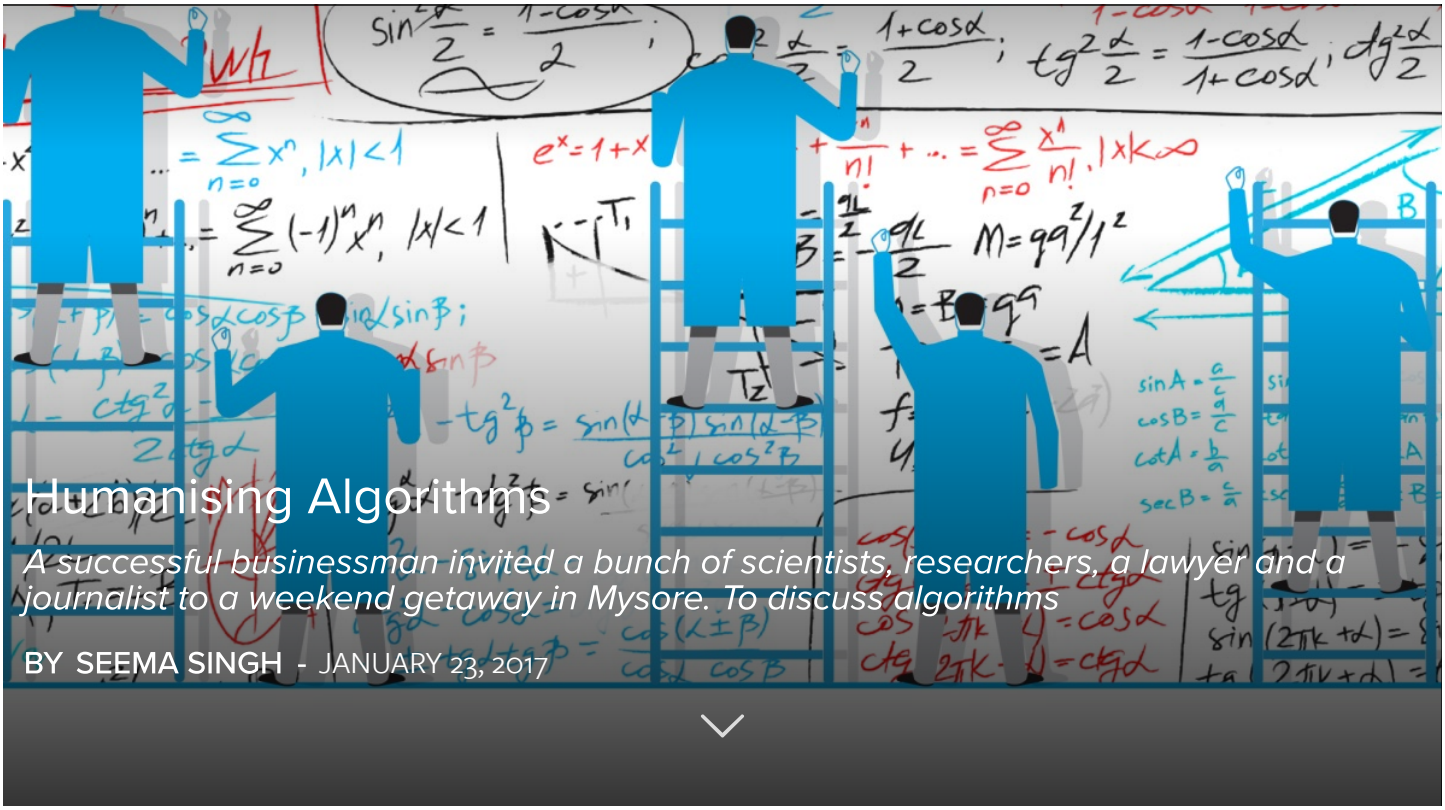

Humanising Algorithms

A successful businessman invited a bunch of scientists, researchers, a lawyer and a journalist to a weekend getaway in Mysore. To discuss algorithms

BY SEEMA SINGH - JANUARY 23, 2017

Give me data and I'll change your world.

Such claims are commonplace in the world of business, governance, healthcare, elections—any walk of life you may think of. The catch words remind us of late John Martin's 'Give me half a tanker of iron and I'll give you an ice age'. One of the catchiest lines ever spoken by an oceanographer, Martin's idea of iron fertilisation of oceans, to trigger plankton blooms that would act as carbon sinks, became popular through the 1990s when global warming was becoming evident. Martin may have said this in half-seriousness, but big data fans, today, are very serious about changing the status quo. Sandy Pentland of Massachusetts Institute of Technology has had ambitions of reinventing human society on a large database platform.

Now, designers of algorithms, which play a critical role in data generation and analysis, are getting serious. Because algorithms are proving to be discriminatory, reinforcing human prejudices, and generally not as objective as they are widely thought to be. And since the march of algorithms is not going to be stopped—India plans to use it for policing—the next best thing to do is to create algorithmic checks and balances that can raise the right flags.

Nearly half of India's 50 or so algorithm designers, theoretical computer scientists to be precise, gathered in Mysore last weekend to brainstorm on such issues. To bring computational lens, a term coined by four Berkeley computer scientists, on other sciences. It doesn't mean applying computer science to a problem, rather it's a way of

looking at the processes of nature (or other disciplines) and giving them a new algorithmic perspective.

Nisheeth Vishnoi is leading the charge. His own work is a prime example of such an approach. Among other things, he has shown a connection between the dynamical systems in signal processing and the single-celled organism, slime mould. Currently, at the Swiss Ècole Polytechnique Fédérale De Lausanne, Vishnoi is exploring if more computer scientists in India could use this lens and if a new Institute of Computation and Human Sciences would make sense.

"We don't want to do big data. We want to understand how we are generating the data. Unless we understand what is special about the data, we don't have any hope of making leapfrog progress," says Vishnoi.

It's no surprise then that Kris Gopalakrishnan, Infosys co-founder, who has supported a few computational science research initiatives in India, has shown deep interest in backing Vishnoi's idea. At the Infosys Leadership Institute in Mysore, Gopalakrishnan kicked off the idea exploration with a fairly broad and ambitious call.

"Computer science is at crossroads. It needs new tools, new algorithms, new software... to create better ways of using a computer or even making better computers. We need new kinds of computing, a new architecture to solve problems of society. Let us think without constraints, everything is possible. It could be in physical groups, virtual groups, a centre within an institute or a new institute itself."

Voilá! Those scientists did think without constraints, for full 36 hours.

How do we save humanity from the tyranny of algorithms? Can we embed ethics in algorithms? Can we create boredom in Artificial Intelligence? To the extent that algorithms are governing our lives, could there be a Constitution of Algorithms in future?

And in all this, where exactly should we place the human?

Putting the human in the loop

In order to get population-wide insight into data, it is imperative that Indian researchers develop the ontology. In simple words, a vocabulary that defines a set of data and their structure so that others, including software agents, can

use it. The right way to do it, suggested Ravi Kothari of IBM Research, is to intervene right at the source, where data is being sensed. Add small chips in the flow of data to enrich it; add descriptors so that live stream data is sufficiently tagged.

"While we have a visual cortex and olfactory system, that is not where complex analytics happens in the brain," said Kothari. In the human brain, the most complex analytics happens in the hippocampus, the place where the fusion of information, from multiple sensing organs, takes place. "Part of the reason, people believe, is that there are a sufficient number of descriptors that are being attached to the information as it comes here."

"But the more descriptors you attach, the more specifically you can profile people who are privacy-destructive in nature," intervened Rahul Matthan, a partner at the law firm Trilegal.

That was the first red flag of the day.

But humans not only make the data more identifiable, they can help improve the machine-based inferencing too. Most statistical methods are not very good with little data. Which is why we hear the near universal cry of 'give me data'. Large data sets are needed to train the learning algorithms so that they can make relevant inferences. Humans, on the other hand, can do well with very little data. We've all made an 'impression in the *first* meeting'.

The discussion wore on. By afternoon, all agreed that we are going to have algorithms, which no one understands. While deep learning may be throwing results, which people are happy with, no one really knows what happens in those 10-11 layers of algorithms.

"For instance, in image recognition, deep learning is doing very well but people probably understand the first layer and maybe the last layer but what happens in the [several] middle layers nobody knows. People are now trying to understand why machine learning gives the [good] results that it gives," Somenath Biswas of IIT-Kanpur later explained.

If so much is opaque in algorithms, can we then bring the human in the loop? Can the human 'inductive' logic either supplement the machine-based inferencing or override it?

"I am wary of bringing humans. Look at the US elections and the wisdom of the crowd. Better to keep humans out of the loop because they bring all kinds of biases," interjected Jayant Haritsa of the Indian Institute of Science.

“You are questioning whether democracy is good or not. You can’t. People chose their president,” argued Vishnoi.

Then should we trust the experts more than the crowd? “They have been getting it wrong, too. Look at the [US] elections result,” suggested another academic.

Keeping the human out of the loop

> If humans are introducing biases and discrimination, then they should be kept out of it, right?

Easier said than done. At the National Intelligence Grid (NATGRID) in Delhi, officials have a challenge, which in computer science is termed as ‘entity resolution’. That is, identifying an individual from tons of data that different agencies collect. Can you reconstruct a digital entity, say an individual or an organisation, from a plethora of digital footprints?

‘We want to embed privacy with security. We have humongous transactional data available but I don’t want anybody to access that. Even if a person from NATGRID is sitting there, I don’t want her to access the data. Can we encrypt the entire data and then let algorithm work on it [for identification]?” said a senior Home Ministry official, also a computer science alumnus from one of the IITs.

Homomorphic encryption (or computing on encrypted data), suggested one. But it's still at the research level, countered another.

So, dig some more into theory.

"In theoretical cryptology, there's a new thing called Zero Knowledge Protocol where you get the result and yet you don't have access to more than what you need to know," said Susmita Sur-Kolay from the Indian Statistical Institute in Kolkata. That's still pretty far out in future, one would say.

While a few vendors have partial solutions but they don't work in the Indian (or even Chinese) context because data in these regions is very noisy. A person may book his ticket in one name; reserve his hotel room using an initial; may have made his ration card under an incorrectly spelt name; stays in a neighbourhood under a nick or pseudo name; and so on.

All of which gives scope to develop fundamentally new algorithmic methods. "It's possible that many of these problems are solved by the National Security Agency [in the US] but there's no way to know what innovations are going on," said Vishnoi.

However, a universally hard problem is keeping bias out of algorithms. Especially so because one person's bias is another person's fact. Fairness is not so much an ethical question as it is a legal question. For instance, why does IndiGo airline have an all-female cabin crew? Bias? No, it's a business decision that makes the aircraft lighter and hence saves the company in fuel costs. Facts are contextual and time-dependent.

The discussion got intense. What if an algorithm gives false positives? Many times, people who develop algorithms don't disclose what parameters they did not use while defining their facts or biases.

It was towards the end of the day, and theorems were flying across the room. As Naveen Garg, from IIT Delhi, got up to present his group's idea on handling bias, he was candid: "From the way I am talking, you can make out I don't understand it very well. It's so complex, it requires synthesis of different things, social sciences, law... Is there a notion of a self-healing algorithm, which adjusts itself to the changing facts over time and modifies itself? We need a new theory to define some of these things."

Do algorithms perpetuate the biases of humans? And are humans feeding into the algorithm through the training data?

"The causality is not clear—whether data is driving human behaviour or human behaviour is driving the data. Just because something is prevalent right now, doesn't mean that it should have been," said Vishnoi. When a human makes a biased decision, you can take it to court, can we do that for algorithms? "This is what we want to crack. This should not be an excuse for humans to perpetuate the bias."

It was time for Gopalakrishnan to intervene. "Science does not discriminate, people do. Let's just build good algorithms."

'Science is organised unpredictability'

This statement from Freeman Dyson was in action at the workshop. Computer scientists swung from Turing-award winning ideas to something as humdrum as getting data from public utilities, say railways or state transport departments, and allowing engineering students in 6000 plus colleges to try their hands on data science.

In the age of applications, and in a country where people think computer science is synonymous with programming, theorists are almost an invisible community. But with the level of abstraction the theoreticians work at, they probably have the best analytical tools to solve some of the confounding problems of our times. Be it in privacy or security, healthcare or education.

I got a glimpse of that in Mysore. From what initially looked like a list of popular tech jargons—Internet of Things, Privacy Preserving Analytics, Entity Resolution, Dreaming and Boredom—emerged well-defined grand challenges. They became grander because Gopalakrishnan encouraged people to think big. For instance, a 'Heat map of infectious diseases' proposed by Vijay Chandru and 'Microbial signatures in disease and health', by the virologist Shahid Jameel, became two interconnected mega ideas that could possibly change the way we look at disease and health in India.

From the group that was brainstorming on 'entity resolution', came the idea of SPIDER — Secure Personal Identification for *Desh Raksha*. (The two Hindi words have less to do with nationalism and more to do with the multilingual nature of the problem.) "If humans can piece together the information, so can an algorithm. We'll find that algorithm," said Pushpak Bhattacharya, a Natural Language Processing expert from IIT-Bombay who is now the director of IIT-Patna.

The computational lens will enrich theory, thinks Biswas. “How did mathematics develop? In trying to solve the problems of astronomy...The [theoretical computer science] community in India is starved of stimulation. These hard problems will energise it.”

What this boot camp will finally result in or how the five big ideas will be taken up, we don’t know. (Gopalakrishnan refused to speak beyond what he had said in the workshop). But what is apparent though is that the community is ‘becoming more outward’, willing to touch the reality threshold.

PS: I was invited to the workshop as a media professional, to represent the society voice as it were. It’s a different matter altogether that I ended up writing about it.

Seema Singh

Writer and Editor