Big data in the physics lab

Methods of the market place are being pressed in for bulk handling of research data

S ANANTHANARAYANAN

The LHC at Cern near Geneva

Research in the frontiers of physics works with very high energy reactions, which generate quantities of data that we have not encountered before. Alexander Radovic, Mike Williams, David Rousseau, Michael Kagan, Daniele Bonacorsi, Alexander Himmel, Adam Aurisano, Kazuhiro Terao and Taritree Wongjirad, from Universities in France, Italy and the US, in the journal, *Nature*, review the challenges and opportunities in using machine learning for dealing with this avalanche of information.

Physics, for all its success, remains an incomplete study. The standard model of particle physics, which describes the known fundamental particles, is unprecedented in the accuracy of its predictions at the very small scale. But this model does not deal with the force of gravity. Gravity has negligible effect at the very fine scale, because masses are so small and electrical or nuclear forces are so much stronger but the theory is not suited for large masses and distances, where the force of gravity dominates.

Even at this small scale, there are loose ends, which need to be tied up. The properties of the neutrino, a very light and uncharged particle and very difficult to encounter, or the Higgs Boson — the particle implicated in the inertia of matter and difficult to create — have not been observed and documented. To complete the understanding of even the standard model, we need to see how matter behaves at distances that are even smaller than the closest we have seen so far. As all matter is repulsive at an extremely short distance, we need to create very high energies to bring particles close enough.

This is the motivation for the high-energy accelerators that have been constructed over the decades and the greatest of them all is the Large Hadron Collider at the European Organization for Nuclear Research — more commonly known by its acronym, Cern — near Geneva. The LHC, over its course of 27 km, accelerates protons to nearly the speed of light, so that their collision would have the energy required to create new, massive particles, like the Higgs. While the Higgs itself is a rare event, there are billions of other products of the collisions, which need to be detected and recorded.

The authors of the *Nature* article write that the arrays of sensors in the LHC contain some 200 million detection elements and the data they produce, after drastic data reduction and compression is as much every hour as Google handles in a year!

It is not possible to process such quantities of data in the ordinary way. An early large data handling challenge faced was while mapping the 3.3 billion base pairs of the human genome. One of the methods used was to press huge numbers of computers, worldwide into action. That was through a programme in which all computers connected to the Internet could participate and contribute their idle time (between keystrokes for instance) to the effort.

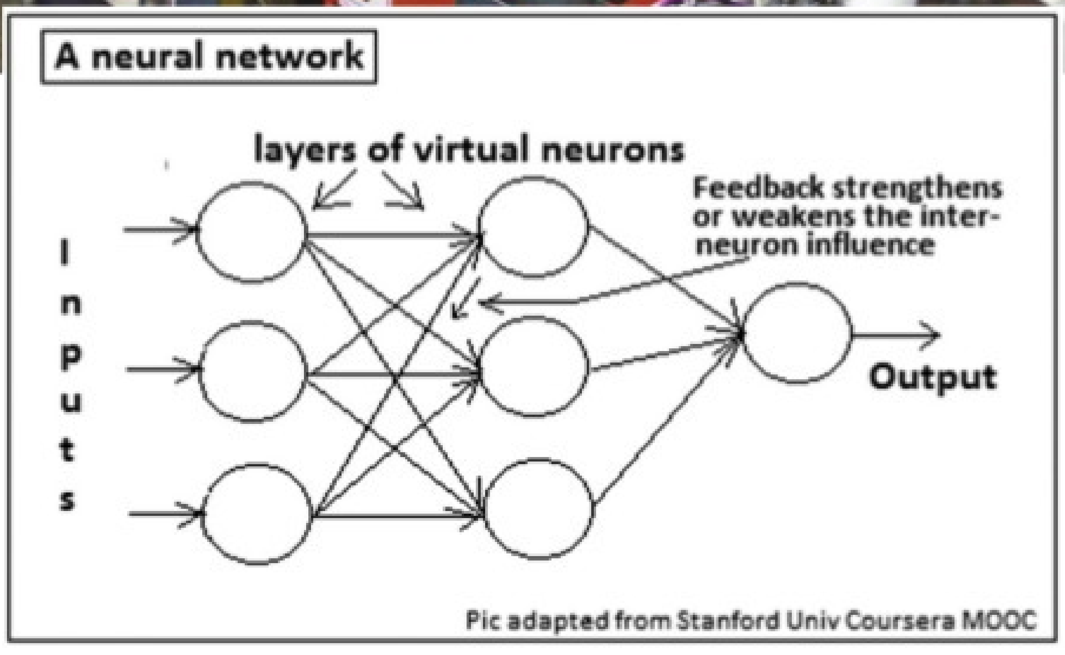
Such methods, however, would not be effective with the data that is generated by the LHC. Even before the data is processed, the *Nature* paper says, the data coming in is filtered so that only one item out of 100,000 is retained. While machine-learning methods are used to handle the data finally used, even the hardware that does the initial filtering uses machine language routines, the paper says.

Machine learning consists of computer and statistical techniques to search for patterns or significant trends within massive data, where conventional methods of data analysis are not feasible. In simple applications, a set of known data is analysed to find a mathematical formula that fits their distribution. The formula is then tested on some known examples and refined, so that it makes correct predictions with unknown data too.

The technique can then be used to devise marketing strategies, weather forecasts and automated clinical diagnoses. While powerful computers were able to process huge data and perform well, without taking too long over it, it was realised there were situations where the animal brain could do better. The animal brain does not use the linear method, of fully analysing data, of the conventional computer, but "trains itself" to interpret significant data elements and trim its responses on the basis of how good the predictions are.

Computer programmes were hence written to simulate the animal brain, in the form of "neural networks" or computers that behave like nerve cells. In a simple instance of recognising just one feature, it could be presented to a single virtual neuron. The neuron responds at random from a set of choices. If the answer is correct, there is feedback that adds to the probability of that response, and if the answer is wrong, the feedback lowers the probability.

We can see that this device would soon learn, through a random process, to consistently make the correct response. A brace of artificial neurons that send responses to another set of neurons, and so on, could deal with several inputs with greater complexity. A network like this could learn to identify an image as being that of a car or a pedestrian, for instance, and if a pedestrian, whether it is a man or woman!



Pic adapted from Stanford Univ Coursera MOOC

This architecture is now adapted in the LHC to make out which data to keep for further analysis and what to discard. The paper says that suitable algorithms, and neural networks, have been developed for satisfying "the stringent robustness requirements of a system that makes irreversible decisions (like discard of data)."

The paper says that the machine learning methods have enabled rapid processing of data that would have taken many years otherwise. Apart from the need for speed, the algorithms used need to be adapted to the specific signatures that are being looked for. The decay products, for instance, that could signal the very rare Higgs Boson event, among others. The paper says that the data handling demands will only increase, as the data from the LHC will grow by an order of magnitudes within a decade, "resulting in much higher data rates and even more complex events to disentangle." The machine learning community is hence at work to discover new and more powerful techniques, the paper says.

Discoveries in physics have traditionally been data based. A turning point in the history of science could be when Copernicus firmly established that Earth and the planets went around the Sun and not that the Sun and planets went around the Earth. Copernicus came to this epochal conclusion entirely on the basis of massive astronomical data collected by his predecessor, Tycho Brahe. The data, collected with rudimentary instrumentation, was voluminous and the analysis was painstaking.

In contrast, we are now using massively sophisticated computing power and much more expensive data acquiring methods. While the puzzle solved by Tycho Brahe and Copernicus was basic and changed the course of science, the problem now being looked at is of a complexity and character that could not have been imagined even a century ago.

The writer can be contacted at response@simplescience.in

PLUS POINTS

Moderation is key

Teetotallers in middle age are at greater risk of developing dementia than those who drink moderate amounts, with benefits particularly apparent in wine drinkers, a new study has found.

Researchers found abstinence was associated with a 45 per cent increase in the chances of getting dementia by early old age, compared to those who drank within recommended limits — up to a bottle and a half of wine a week.

People who drank above the 14 unit guideline were also at increased risk, the team from University College London and French institute for health, Inserm, found. Their risk of developing dementia increased incrementally the more alcohol they were consuming.

"We show that both long term alcohol abstinence and excessive alcohol consumption may increase the risk of dementia," the authors of the study, published in the *British Medical Journal*, wrote. "Given the number of people living with dementia is expected to triple by 2050 and the absence of a cure, prevention is key."

"Wine, in addition to alcohol, contains polyphenolic compounds, which have been associated with neuro-protective effects on both neuro-degenerative and vascular pathways, and with cardioprotective effects through inflammation reduction, inhibition of platelet aggregation, and alteration of lipid profile," said Sevil Yasar from John Hopkins School of Medicine in Baltimore, US, who was not involved in the study.

Among abstainers, the study also found increased risks of diabetes and cardiovascular disease. These are both conditions, which could contribute to dementia, a collective term for the loss of memory, thinking and other cognitive functions.

The authors said that they would not encourage anyone currently abstaining to change their habits.

The Independent

Engineered rice

Rice plants engineered to have fewer stomata — tiny openings used for gas exchange — are more tolerant to drought and resilient to future climate change, a new study has revealed.

Scientists from the University of Sheffield, UK, have discovered that engineering a high-yielding rice cultivar to have reduced stomatal density, helps the crop to conserve water and survive high temperatures and drought. Much of humanity relies on rice as a food source, but rice cultivation is particularly water intensive — using an estimated 2,500 litres of water per kilogram of rice produced. However, almost half of the global rice crop derives from rain-fed agricultural systems where drought and high temperatures are predicted to become more frequent and damaging under climate change.

Like most plants, rice uses microscopic pores called stomata to regulate carbon dioxide uptake for photosynthesis, along with the release of water vapour via transpiration. When water is plentiful, stomatal opening also permits regulation of plant temperature by evaporative cooling. Under water-limiting drought conditions, stomatal closure normally slows down water loss. Low stomatal density rice conserves its water better under drought, and so has more water left to cool itself when necessary.

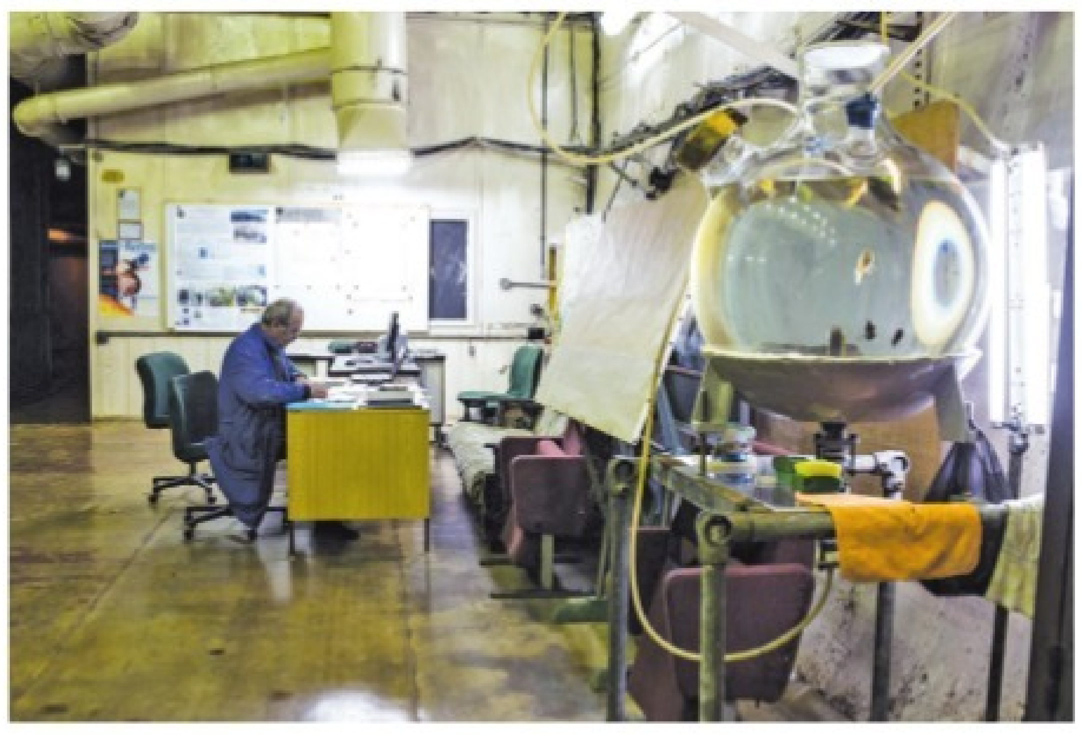
Robert Caine, research associate from the University of Sheffield's department of molecular biology and biotechnology and first author of the study, said, "We found that the engineered rice crops gave equivalent or even improved yields, which means it could have a massive impact on our future food security, which is threatened by climate change."

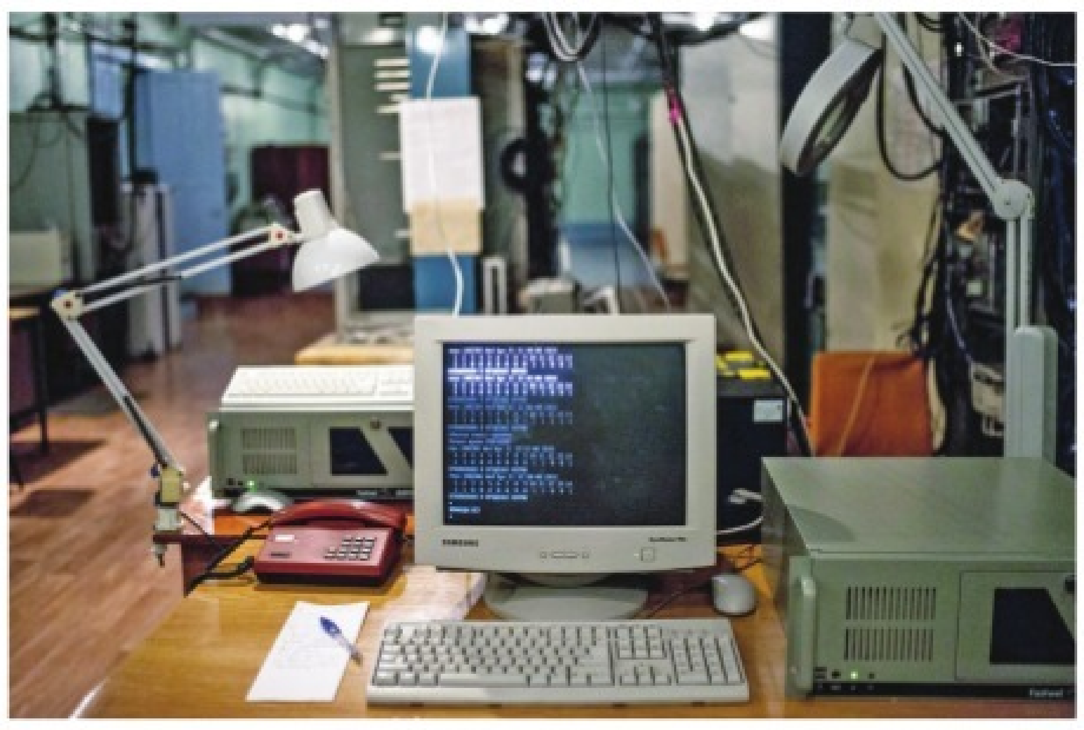
The new study, published recently in *New Phytologist* and conducted in collaboration with the International Rice Research Institute in the Philippines, found low stomatal density rice lines used just 60 per cent of the normal amount of water. When grown at elevated atmospheric carbon dioxide levels, the low stomatal density rice plants were able to survive drought and high temperature (40 degrees Celsius) for longer than unaltered plants. To view full paper, check out https://nph.onlinelibrary.wiley.com/doi/full/10.1111/nph.15344

Engaged in a surreal pursuit

Deep in the Caucasus Mountains, scientists are trying to listen to quantum whispers about the nature of reality

DENNIS OVERBYE





Just over the border from Georgia, in the Caucasus Mountains of southern Russia, lies the small town of Neytrino. For the last half a century, its main business has been the study of the tiniest insubstantial bit of matter in the universe, an ephemeral fly-by-night subatomic particle called the neutrino.

This is the home of the Baksan Neutrino Observatory, a warren of tunnels and laboratories burrowed two miles into a mountain, sheltered from the outside universe and cosmic rays underneath 12,000 feet of rock. There, vats of liquid wait to record the flight of neutrinos from the centre of the sun, from exploding stars, atomic reactors and the Big Bang itself, carrying messages through time.

Neutrinos are the ghost riders of the cosmos, mostly impervious to the forces, like electromagnetism, with which other denizens of nature interact. Neutrinos cruise unmolested through rocks, the earth and even our bodies. In the words of a famous poem by John Updike, they "insult the stallion in his stall". The most delicate measurements so far indicate that an individual neutrino weighs less than one millionth of what an electron weighs.

Baksan is not the only place dedicated to their surreal pursuit. The men and women of Neytrino share an underground union with scientists scattered around the world in equally deep places — the Sanford Underground Research Facility in the former Homestake gold mine in Lead, South Dakota; the Gran Sasso national laboratory, beneath the mountain of that name in Italy; the Sudbury Neutrino observatory in Ontario, Canada; the Super-Kamiokande, deep within Mount Ikeno, Japan; and IceCube, an array of detectors buried in ice at the South Pole. All of them are trying to listen to quantum whispers about the nature of reality.

One of Baksan's biggest claims to fame to date is having caught neutrinos emitted by thermonuclear reactions in the centre of the sun in nearly 60 tonnes of liquid gallium. The experiment, called Sage (Soviet-American Gallium Experiment), proved that scientists actually do know what powers our favourite star, source of our life and light. Since the fall of the Soviet Union, the scientists in Baksan have had to fend off both thieves and the Russian government to keep their gallium, an element that goes for some US $500 a kilogram.

Physicists know that neutrinos come in at least three flavours, known as electron, muon and tau neutrinos, depending on their subatomic origin. To add to the confusion, neutrinos have a kind of quantum superpower -- they can molt from one type to another, sort of like a prison-breaker changing clothes during the escape. An electron neutrino, say, can emerge from a nuclear reactor in one place and appear in a detector somewhere else as a muon neutrino. This complicates the cosmic accounting of these creatures.

Physicists are arguing intensely these days over whether there is evidence for a fourth type, called sterile neutrinos. That is the object of a new experiment called Best (Baksan Experiment on Sterile Transitions), now underway in the rusty Baksan tunnels.

Although neutrinos are the lightest and flimsiest and perhaps most fickle particles of the universe, they are also the most numerous, outnumbering the protons and electrons that make up us and ordinary matter by a billion to one. And so neutrinos contribute about as much mass to the universe as the visible stars. An extra population of neutrinos discovered by scientists in a cave in the Caucasus would affect basic calculations of the expansion of the universe.

The discovery this month of a high-energy neutrino from a far distant galaxy passing through the IceCube detector at the South Pole elicited headlines around the world.

Meanwhile, unaware that they are being harassed by extraterrestrial visitors, horses graze outside Baksan, and life goes on, whether we understand it or not.

Courtesy the New York Times

The Independent