SUMMARY OF DISCUSSION

CAMRI Executive Roundtable Luncheon Series
"Quantum Technologies, AI, Big Data, & the Internet of Things"
By
Dr Steven E. Koonin
Director, Center for Urban Science and Progress, NYU
Friday, 5 January 2018

Advances in computing hardware and digitalization of records is resulting in a data revolution. Big data, artificial intelligence (AI), and quantum technologies have attracted unprecedented attention from developers in various industries, as well as researchers of different disciplines. The so-called data revolution has the potential to change people's life from every perspective, yet society must adapt and develop to meet the challenges, which are not only from the introduction of novel forms of data (i.e., Big Data), but also from the effective approaches to utilizing the data.

Within this context, the Centre for Asset Management Research & Investments (CAMRI) invited Dr. Steven E. Koonin of NYU to speak on "Quantum Technologies, AI, Big Data, & the Internet of Things", where his goal was to:

- review the progress in processing power, memory, storage, and sensors during the past few decades;
- understand the characteristics, uses, and issues of Big Data;
- identify the approaches in artificial intelligence (AI) technologies and the applications in urban planning and real life; and
- introduce a non-classical computational methodology – quantum technology and computation, and discuss its importance and future challenges.

Dr. Koonin is the founding director of NYU's Center for Urban Science and Progress since April 2012. Prior to his NYU appointment, Dr. Koonin served as the second Under Secretary for Science at the U.S. Department of Energy from May 2009 through November 2011. The Center of Urban Science and Progress' aim is to develop, demonstrate, and apply technologies to acquire, integrate, and analyze urban data, e.g., records data, sensor data on infrastructure, environment, and people, etc. The working group assembled from the industry and academy had a robust roundtable discussion on some critical issues and challenges that have arisen from the data revolution, AI and the Internet of Things.

Dr. Koonin started the discussion by posing some questions:

- What if we could measure "everything" in the natural world, the built world, and even human behavior?
- Which variables would we want to measure and on what scale?

- Implementation-wise, how do we go about data acquisition, transmission, storage, integration, and analysis?
- What would we do with the data? How do we go from data to information, to knowledge, to understanding, and, finally, prediction?

Progress in hardware, specifically processing power, has been spectacular over the last few decades. On average, processors are 10 times faster than those from a decade ago, and CPUs have been a major contributor to that progress. On top of speed gains, we are also getting very good at aggregating these processors into large high-performance systems. The average performance of the top 500 computing systems has improved by about 1,000 times over the last decade. In the last 5 years, Chinese computing systems have started to appear in the Top 10. Researchers globally are clearly pushing hard to improve high performance computing, as well as system usage and software visualization.

At the same time, computing memory and storage are becoming exponentially cheaper - the cost per megabyte in US dollar terms is about 100 times cheaper than a decade ago. Sensors embedded in machinery are also becoming cheaper and ubiquitous. It is now possible to place sensors everywhere, from vehicles to personal devices, and even microphones and cameras. This is leading to a growing number of connected devices. Some estimate that by 2020, 30-50 billion devices will be connected to the internet (4-7 devices for every human on earth). To sum up, it is clear that the "instrumentation of everything" is underway at the industrial level, the city level, and the consumer level.

The progression in computing hardware, sensors, and communications has made it possible for the emergence of a new form of data: Big Data. Big Data is an extraordinarily active area of research, not only in academia but also in the commercial world. There are hundreds of firms using analytics as well as building the data infrastructure. In contrast to classical data from 20 years ago, Big Data can be characterized by the 5 V's – Volume, Variety, Velocity, Veracity, and Value. First, data now is much more voluminous; it often can hardly fit in a spreadsheet or even a laptop. Second, data can be both structured – for example, financial, administrative, and transactional data – and unstructured – for example, text, audio, images, video, geometric, and genomic. Third, the rate at which data is generated can be extraordinarily fast, in Gigabits/second in some cases. Fourth, data is noisy, incomplete, and ambiguous. Fifth, we must discern the value of data, i.e., "not everything you can know is interesting."

In order to illustrate the large volumes involved in Big Data, Dr. Koonin took sensor data from a cross-country flight, and the personal and transactional data from a city, as two examples. For an airplane that flies cross-country within the US, each engine produces 20 terabytes of information every hour; and there are 28,537 commercial flights in the sky in the United States on any given day. Using a city with 3 million people as another example, within a radius of 10 km in a dense urban area ($10^4/km^2$), they live in 1 million households, work in 0.3 million businesses, and own or use 0.5 million vehicles. Updates of personal data, such as physical location, physiology and sociology data generate 20 bytes every 5 seconds, which is equivalent to 12 megabytes per second. Transactional data, such as credit card payments, are less frequent at around 20 megabytes per second. This amounts to around 600 terabytes of data per year for the city. Yet, storage costs are less than USD100,000.

There are wide uses of Big Data, categorized as resource allocation, outlier identification, predictive models, and effects of intervention. One example given by Dr. Koonin is using taxis as sensors in New York City. Taxis are sensors that can provide unprecedented insight into city life, including economic activity, human behavior, and mobility patterns; 500 thousand taxi trips are made per day in NYC, which equates to 180 million trips per year. Big Data allows us to answer questions such as: a) what is the average trip time from midtown to the airport during weekdays? b) How does taxi fleet activity vary during weekdays? c) How is taxi activity in midtown affected during a Presidential visit? d) How do movement patterns change during Sundays? e) Which are the popular night spots?

However, as pointed out by Dr. Koonin, there are also critical challenges: a) storage, transmission, compression, and feature extraction, b) data standards and quality, c) linking of records or combination of multiple database, d) streaming or real-time analytics, e) security and privacy, f) legalities of data access, and g) infusing domain knowledge.

A critical approach to Big Data has to be applied in order to impact real life. Artificial intelligence (AI) technologies are well-known, often discussed, and are driving the next wave of innovation. How do we even define artificial intelligence? On the one hand, the artificial intelligence (AI) that people usually talk about refers to the task-focused artificial intelligence, which is making rapid progress in recent years. Task-focused artificial intelligence, which includes machine learning, machine vision, machine listening, natural language processing, text analysis, social media analysis, biological and health analysis, and robotics, is making rapid progress. AI engines can equal or exceed human performance in games (Chess in 1997, Jeopardy in 2011, Atari video games in 2015, and Go in 2016), and in image understanding. It is an essential tool in exploiting big data. Artificial general intelligence (AGI), on the other hand, is a small research field within the broader field of AI, which goal is to “perform any intellectual task a human can”. Despite catastrophist warnings from Stephen Hawking, Elon Musk, and other thought leaders, Dr Koonin thinks it will remain a topic for futurists and science fiction writers for the foreseeable future.

Deep Learning, embodied in a Deep Neural Net (DNN), has come to dominate AI in the past decade. For example, a deep Conventional Neural Net (CNN) is best for processing images. Such deep learning is enabled by a) CPUs that are fast and cheap computing engines, b) large labeled data sets (images, sounds, and records) for training, c) algorithms to efficiently train some types of nets, d) usable software tools to build and train nets, and e) demonstrated success in multiple tasks.

While neural nets are inspired by the human brain, there are, however, some design issues with the DNN: a) how many layers and how many nodes should there be in each net? b) is it feed-forward or recurrent? c) how connected do the nodes have to be? d) how to train or adjust the weights and thresholds?

One of the examples given by Dr. Koonin, in order to help understand how capable a Deep Neural Net (DNN) can be, is that the machine can understand four different aspects of an image: (a) image segmentation – identifying the pixels that constitute Person A, (b) pose estimation – using a knowledge of human anatomy to construct a likely three-dimensional pose of Person B, (c) associating groups of objects, as that Person C is playing an accordion, and (d) recognizing and constructing 3-D models of partially hidden objects, i.e. the bench.

Some other examples of applying Deep Learning in daily life include: a) DNN for pedestrian counts, b) phonogram that turns audio into an image to then be understood, and c) machine listening for real-time sound identification.

Dr. Koonin also provided eight rules of applying Deep Learning in practice:

- Use deep (or very deep, where possible) neural nets.
- Use convolutional nets, even if you don't know why (that is, even if the underlying problem is not translation invariant).
- Adopt flat numerical data representations, where the input is a vector of reals and the internal representation (for a DNN, the activations) is an even larger number of reals.
- Avoid the use of more complicated data structures. The model will discover any necessary structure in the data from its flat representation.
- Train with big (*really* big) data.
- Do not load on model assumptions, but rather learn everything from the data—that is where the truth lies. With enough data, it is more efficient to let the DNN discover them on its own.
- An approximate answer is usually good enough.
- When it works, it is not always necessary to understand why or how.

In the meanwhile, there are still challenges in using Deep Learning: (a) large labeled training sets are required, which are difficult to obtain, (b) construction and training of a net is *craft*, not *engineering,* (c) unless architected, initialized, and trained identically, two nets might have comparable macro performance, but will differ in detail (the "close calls"), and (d) a net's "reasoning process" cannot be explained easily, if at all; it is difficult to formally control software development and maintenance, difficult to manage performance (especially in the tails), and difficult to judge fairness and bias.

In the final part, the discussion moved to quantum information, which is a new science that combines quantum mechanics and information science with five underlying sub-areas: a) Fundamental physics, b) Quantum cryptography, c) Quantum communication, d) Quantum computation, and e) Quantum metrology. The discussion focused on two areas – Quantum computation and Quantum communication.

Quantum computation contrasts with classical computation in terms of methodology. The basic unit in classical computation is the bit (0 or 1), while the basic unit in the quantum computation is the qubit which is a state of a quantum two-level system and can be simultaneously $|0\rangle$ and $|1\rangle$. Despite its spectacular progress, classical computation is limited. For example, in classical computation, transistors must have more than a few atoms, and some problems are so difficult that they are beyond any foreseeable classical capabilities. Quantum computation can bypass these limits since it carries more information.

Quantum computation has been rising in stature and has promising prospects. D-Wave has a 1,152 qubits optimizing engine, while the state of the art for a general-purpose quantum computer is around 20 qubits. In addition, we expect to have around 1,000 qubits general-purpose machines within 5 years as both big companies and startups are investing in quantum computation. However, there are also challenges in the area of quantum computation: (a) decoherence and imperfections – since interaction with the external environment destroys quantum superposition, coherence must be maintained long enough to perform a calculation, and imperfections can occur in the initial state, quantum gates and final state projection, (b) error correcting codes make quantum computation "fault tolerant", which increases the required number of physical qubits by an order of magnitude, (c) scaling to useful number of qubits, and (d) different algorithms (Shor for factoring, Grover for sorting) and large-scale optimization are needed.

Quantum technology enables secure communications: Photons (light) are transmitted through a fiber or through free space from/to a satellite. The photon polarization states form the qubit and quantum coherence makes the transmission resistant to interception. Importantly, the quantum encryption keys cannot be learned retroactively.

Dr. Koonin ended the discussion with some of his thoughts on Quantum Key Distribution (QKD). The first implementation of QKD is in 1992 and nowadays it can reach a rate of 1 mbps over 50 km fiber or 10 kbps over 100 km fiber. The largest QKD network is in China, which launched its Micius satellite in August, 2016. However, it is still in its early days and we expect more sophisticated systems exploiting more capable quantum resources, including growing range in fiber, growing bit rates, routine satellite links, links to ground, and airborne vehicles.