OVERVIEW of CMSC 828Y: Human-level AI

Spring 2017

AI has been around as a subject of coordinated study for roughly 60 years.

One of its primary and early-stated goals was to come to a deep computationally-based understanding of the nature of the mind, and especially of human intellectual capacity. The idea was that the brain is an enormous information-processing device, honed to its amazing capacities by eons of evolutionary steps, but which could perhaps be mimicked in computers much more quickly.

This was not so much intended to replace humans, or to populate the world with smart machines, as to *understand* the mind far better. People have long been keenly interested in knowing more about ourselves, whether as psychologists or sociologists or biologists or philosophers. Who – or what – are we? This motivated people to study the brain and to form models of cognitive processes at various levels, from examination of the microscopic structure of the brain to high-level abstract computation.

A thumbnail pre-history of AI (learning about how the brain processes info):

1870s-80s: Golgi and Cajal – discovery of neural units (neurons)

1943: McCulloch and Pitts – abstract computational model of neural processing

1949: Hebb – Organization of Behavior (theory of neural learning)

1952: Hodgkin and Huxley: math model of electrochemistry of neural signals

A thumbnail very early history of AI:

1950: Turing – speculations on computing and intelligence

1956: Dartmouth Conference

1958: McCarthy – Programs with common sense (the Advice-Taker)

But over the years, successes in AI technology have had such an impact that nowadays it is sometimes that original aim does not seem much in evidence. Or, put the other way around, sometimes it seems as if that original aim is almost here – that machines are getting better than humans at more and more, and soon we will be outmatched.

But the truth is very different. Or rather, much more complicated.

Yes, machines can be programmed to do some things better than humans – play chess, or win at Jeopardy, and so on; and not just games. But virtually all of these successes are in narrowly constrained domains, where so-called expertise is called for. A human can train and get pretty good in a tightly specified domain, and so can a machine. So far, machines do better than humans in some cases, and in others humans do better.

But humans also are good in another less constrained way: we manage to “muddle through” in the world in general, where we don’t have specific expert training. We deal with new things on a daily – even hourly – basis. We tend to see our errors and either fix them or ask for help or back away gracefully or plunge ahead because they don’t matter, or decide to work at improving our skills for next time.

Especially, we ask for help. That is, we absorb information from our culture (the enormous fund of wisdom all around us, passed on and enhanced over centuries). This includes Google, books, conversations, just watching what others do, and so on. The phrase “commonsense behavior” is sometimes used as a gloss for this. And it is not just memorization of new facts; it crucially involves the ability to understand and draw new conclusions. And this in turn depends on model-building, creating and manipulating internal structures so we can anticipate how a piece of the world is behaving, or how it would behave under imagined conditions.

Is this a combination of billions of independent special cases that have become hard-wired in our brains, hence of enormous complexity and not comprehensible in terms of basic mechanisms? Or are there definite organizing principles that – once we have found them – will put the entire range of human mental life into sharp relief, much as DNA has done for biology?

It is an open question. Many in AI (and elsewhere) think such principles may be found. But we have a long way to go, despite all the amazing successes to date.

Deep learning – the exciting new kid on the AI block – is one example of a technology that grew largely out of neuroscience. And biologically-inspired AI is playing a larger and larger role.

And yet, to quote Sergio Bengio – one of the foremost proponents of deep-learning – “machines (today) are stupid”. I would go further, suggesting that they aren’t even stupid yet. To be stupid, one at the very least has to have beliefs about the world. And this turns out to be totally absent for virtually all of AI, even in the areas of commonsense reasoning, natural-language processing, and image-processing. This may seem like nonsense – but in taking CMSC 828Y you may come to think otherwise. The recent emphasis on bringing AI and robotics together presents a situation where this can no longer be ignored: a reasoning robot has to relate its reasoning to the world it interacts with.

CMSC 828Y will focus on the above issues with regard to achieving human-level intelligence in general: what has been done, what is being done today, and where this may lead. The format will be lectures and discussion; workload will include readings, quizzes, papers and presentations. Topics will range widely, from AI methods and subdisciplines to aspects of cognitive psychology and neuroscience.