About Deep Learning, Intuition and Thinking

by Fabrizio Falchi, (ISTI-CNR)

In recent years, expert intuition has been a hot topic within the discipline of psychology and decision making. The results of this research can help in understanding deep learning; the driving force behind the AI renaissance, which started in 2012.

"Intuition is nothing more and nothing less than recognition" [1], is a famous quote by Herbert Simon, who received the Turing Award in 1975 and the Nobel Prize in 1978. As explained by Daniel Kahneman, another Nobel Prize winner, in his book Thinking, Fast and Slow [2], and during his talk at Google in 2011 [L1]: "There is really no difference between the physician recognising a particular disease from a facial expression and a little child learning, pointing to something and saying doggie. The little child has no idea what the clues are but he just said, he just knows this is a dog without knowing why he knows". These milestones should be used as a guideline to help understanding decision making in recent AI algorithms and thus their transparency.

Most of the recent progress in artificial intelligence (AI) has been on recognition tasks, and this progress has been achieved through the adoption of deep learning (DL) methods. The AI renaissance started in 2012 when a deep neural network, built by Hinton's team, won the ImageNet Large Scale Visual Recognition Challenge. Deep learning methods have been, and still are, the driving force behind this renaissance. Like the little child mentioned by Kahneman, a state-of-the-art deep neural network is able to look at something and say "doggie", without knowing why it knows. In other words, the task of recognition, especially in computer vision, has been solved by DL methods with a form of artificial intuition. And this is not a surprise given that important researchers such as Simon have accepted the equivalence between intuition and recognition.

Even if many people feel a sense of magic talking about DL, the research conducted in recent years has proven that there is no magic at all in intuition and the same holds for DL.

Within the discipline of psychology and decision making, expert intuition has been discussed a lot in recent years, dividing researchers into believers and skeptics. However, after six years of discussion, a believer, Gary Klein, and a skeptic, Daniel Kahneman, wrote an important paper in 2009 whose subtitle was "A failure to disagree". Trying to answer the question When can we trust intuition? they agreed on a set of conditions for trustable intuitive expertise. Among these conditions, the most important ones are:

- an environment that is sufficiently regular to be predictable;
- an opportunity to learn these regularities through prolonged practice.

I believe most of the researchers working on DL would agree that those are also good conditions for the question: When can we trust deep learning? In fact, in order for a DL method to learn, we need a large training set (prolonged practice) and this set must be representative of the application scenario in which the environment must be sufficiently regular to be predictable.

What degree of transparency can we ask for from DL methods? Following the metaphor between DL and intuition, we can look at what Simon said about human recognition capabilities: "we do not have access to the processes that allow us to recognise a familiar object or person". I believe the same is true for DL. Even if we can monitor the flow of information in a deep neural network, we don't understand the "process".

Nevertheless, DL methods can be transparent in some terms: knowledge about the used training set and an in-depth analysis of the statistical outcomes can help in making them trustable for a specific task, in a specific context at a specific time.

As humans should not rely on intuition for all decisions, DL methods should be used as part of more complex AI systems that also involve non-intuitive processes. Kahneman has used the metaphor of two systems in his research about human thinking:

- System 1: fast, automatic frequent, emotional, stereotypic, unconscious.
- System 2: slow, effortful, infrequent, logical, calculating, conscious.

It is not by chance that DeepMind AlphaGo, the program that in 2016 defeated South Korean professional Go player Lee Sedol, combines DL with Monte Carlo tree search. As Michael Wooldridge, chair of the IJCAI Awards Committee said, "AlphaGo achieves what it does through a brilliant combination of classic AI techniques as well as the state-of-the-art machine learning techniques that DeepMind is so closely associated with". Following our metaphor, AlphaGo is a good example of collaboration between System 1 and System 2. AlphaGo uses DL to provide an intuitive estimation of the likelihood that the next stone will be placed in a specific place and of the final outcome of the game given the current status. However, the final decision about where to put a stone is made using the Monte Carlo tree search (AlphaGo System 2). In other words, AlphaGo uses the outcome of artificial intuition implemented using DL methods (its System 1) but takes decisions with logical reasoning (its System 2).

The few examples discussed here show that psychology can help in understanding AI. When Simon was conducting his research, psychology and AI were closely linked. This is a link that we need to revisit.

Link: [L1] https://kwz.me/hdj

References:

[1] H. A. Simon: "What is an 'explanation' of behavior?" Psychological science 3.3, 150-161, 1992.

[2] D. Kahnemann: "Thinking, Fast and Slow", Farrar, Straus and Giroux, 2011.

Please contact:

Fabrizio Falchi, ISTI-CNR, Italy
+39 050 315 29 11
Fabrizio.falchi@cnr.it