# Do we still need models or just more data and compute?

*Max Welling, Amsterdam, April 20 2019*

This question, or versions of it, seems to divide the AI community. And much like Bayesians and Frequentists they hold rather strong polarizing views on the matter. The question seems to come in different flavors: symbolic AI or statistical AI, white box AI or black box AI, model driven or data driven AI, generative or discriminative AI? A recent blog by Rich Sutton adds to the list compute-driven AI versus human-knowledge based AI. The discussion is both fascinating and deeply fundamental. We should all be thinking about these questions.

Let me start by stating that I am a strong believer in the power of computation and its relevance to ML. One of the reasons I have a part-time position at Qualcomm is precisely because I believe one of the fastest ways to make progress in AI is to make specialized hardware for AI computations. I am also a strong proponent of deep learning. Much of my research portfolio is currently based on DL. I think it is the biggest hammer we have produced thus far and I witness its impact every day. (So please, Gary Marcus, do not write after reading this post that I am an opponent of DL).

In the blog by Rich Sutton, "The Bitter Lesson", one can read something of the kind: one should work on scalable methods that can maximally leverage compute and forget about modeling the world. A number of examples are explained to support this claim, namely Deep Blue and AlphaGO who leverage search and learning rather than human strategies, speech recognition, visual object recognition etc. And we can add a few more to the list, melanoma detection and tumor detection, statistical machine translation etc. There is no doubt a trend here that cannot be ignored.

But from Rich's argumentation there is one really important factor missing: besides compute, data is perhaps the more fundamental raw material of machine learning. All the examples above share one crucial property, namely that they are very well, and rather narrowly defined problems where you can either generate your own data (e.g. alphaGO) or have ample data available (e.g. speech). In these regimes data-driven, discriminative, black box methods such as DL shine. We can view this as interpolation problems. The input domain is well delimited, we have sufficient data to cover that input domain and interpolate between the dots. The trouble starts when we need to extrapolate.

From my education in ML I recall one thing most vividly: there are no predictions without assumptions, no generalization without inductive bias. Even data driven methods such as DL use assumptions such as smoothness of the function we try to estimate, translation invariance for CNNs, and a hierarchical organization of concepts. More recent advances have tried to build on these rather weak assumptions by for instance adding more types of symmetries to the convolutions. The most fundamental lesson of ML is the bias-variance tradeoff: when you have sufficient data, you do not need to impose a lot of human generated inductive bias on your model. You can "let the data speak". However, when you do not have sufficient data available you will need to use human-knowledge to fill the gaps. This is precisely what happens when you extrapolate: you enter a new input domain where you have very sparse data and your trained model will start to fail.

Note that in certain RL problems, such as learning to play chess or GO, the input domain is so simple that you can easily build an almost perfect simulator, and hence generate an infinite data set without looking at the real world. In that case the most constraining factor is compute and not data. But a simulator is a model of the world and so I would argue that is a model-based approach to RL with human knowledge embedded into the problem. It's just that the world is very simple in this case.

But now consider the problem is a self-driving car. In this case there is a very long tail of traffic situations that are very rare and therefore do not show up in your dataset. In this case a purely data-driven method that does not try to model the world is doomed in my opinion. And in fact, these "exceptions" are currently still captured by a rule-based system. You need to understand how the physics of the world works, as well as the psychology and sociology of the people that populate it, perhaps all encapsulated in a simulator, in order to (learn to) plan in that world and in order to be able to face new, never encountered situations. And the problem is that there are simply too many new situations to hope we can get sufficient data for all of them.

But there is hope. The key insight is that the world operates in the "forward, generative, causal direction". It's the direction in which events cause other events to happen and get recorded on our sensors. We need remarkably few parameters to describe this world: the laws of physics are surprisingly compact to encode. This world is organized modularly, consisting of factors and actors that can be approximately modelled independently. In this world one event causes another event according to the stable laws of physics.

Generative models are far better in generalization to new unseen domains. Causality allows us to easily transfer predictors from one domain to the next: accidents are correlated with black cars in the Netherlands but perhaps with red cars in the US. Using color as a predictor does not generalize, but a causal factor such as male testosterone levels will generalize very well. Generative models allow us to learn from a single example because we can embed that example in an ocean of background knowledge. It would be silly not to exploit the simplicity of the generative direction.

Humans have a remarkable ability to simulate counterfactual worlds that will never be but can exist in our minds. We can imagine the consequences of our actions by simulating the world as it unfolds under that action, or we can derive what caused current events by simulating possible worlds that may have led to it. Clearly, this ability depends on our intuitive understanding of physics and/or psychology.

Discriminative methods operate in the reverse direction. They are mappings from observations directly back to possible causes: predict the word that cause this sound wave in my microphone, predict the object that causes these pixel values in my camera. In this direction things are not simple but everything is highly entangled and complicated. We first need to disentangle the input to make meaningful predictions. But, in narrowly defined domains, these methods work really well because they learn the map that we want to use for predictions. In contrast, inverting a generative model is often intractable and computationally demanding.

It seems we need to find the middle ground. For narrowly defined domains with enough data or a really accurate simulator, you can train a discriminative model and do very well. But when you need to generalize to new domains, i.e. extrapolate away from the data, you will need a generative model. The generative model will be based on certain assumptions about the world, but is expected to generalize a lot better. Moreover, it can be trained and/or finetuned using unsupervised learning without the need for labels. As you are collecting more (labeled) data in the new domain you can slowly replace the inverse generative model with a discriminative model.

Note that I am not claiming that rule based, symbolic AI is the same as building generative models. Knowledge graphs and other structures such as probabilistic variants of those are only one way to inject human knowledge to a model. (Causal) graphical models, and physics-based simulators are other examples. These can all be useful for different tasks.

In conclusion I would say if we ever want to solve Artificial General Intelligence (AGI) then we will need model based RL. And perhaps that model can be learned using unsupervised learning with very little prior human knowledge and with lots of computational resources. In that sense I agree with Rich's bitter lesson. But we cannot answer the question of whether we need human designed models without talking about the availability of data.