Model-Based Reinforcement Learning and Acting for Reasons

Patrick Butlin
King's College London

In an influential recent paper, Lake and colleagues (2017) defend a distinction between *model-building* and *pattern recognition* algorithms, and argue that the use of the former kind will be crucial in the development of artificial intelligence with human-like capacities. While Lake and colleagues focus on measurable aspects of intelligence, we may also consider the role of cognitive models in grounding the applicability of mentalistic concepts such as knowledge, understanding, desire, reasoning and intelligence itself. Here I will consider the proposal that model-based reinforcement learning (RL) is necessary and sufficient – at least against the right background of other capacities – for the capacity to act for reasons.

Model-based RL is distinguished from another form, model-free RL, by the use of representations of learnt action-outcome contingencies, which can be used to anticipate the likely consequences of possible actions (sometimes called 'R-O learning'; Sutton & Barto forthcoming). These can be used together with representations of the values of outcomes in expected utility calculations. In contrast, model-free RL involves learning the value of performing actions under given environmental conditions ('S-R learning'). Meanwhile, to act for a reason is to act because one takes some consideration – a feature of one's current situation – to count in favour of so acting. For example, during my recent holiday in Italy I visited Ostia Antica, because I believed that I would find an outstandingly well-preserved ancient Roman town there; this was one of my reasons for doing so. Actions done for reasons can be contrasted with both actions which are not done for reasons, such as habitual skin-picking, and events which are not exercises of agency at all, such as when my computer automatically locks after a period of inactivity.

The proposal in question is attractive for several reasons. Most fundamentally, according to the highly influential Humean Theory of Motivation, acting for a reason involves desiring some outcome O, and acting in a way that one believes will make O more likely (Smith 1987). The desires and instrumental beliefs on this picture may be taken to correspond to the representations of outcome values and action-outcome contingencies used in model-based RL. Philosophers who deny the Humean Theory tend to argue that the place of desires in practical reasoning can equally well be taken by evaluative or normative beliefs, and this objection does not affect the congruence of the philosophical picture with model-based RL. Furthermore, human neural systems have been interpreted as implementing model-based RL

(Daw et al. 2005), and because the various representations used in action selection by model-based RL systems are learnt, they can be more plausibly identified as the system's own reasons, rather than those of a programmer (Dretske 1993).

The proposal also fits into a broader picture emphasising connections between anticipation, the use of cognitive models, and human-like cognition which is suggested by Lake and colleagues' work, as well as by more widespread intellectual trends. Many theorists, from a variety of disciplines, have emphasised the importance of the use of models of the body and of aspects of the environment in cognition (e.g. Craik 1943, Rosen 1985; but the concept of a model is continually invoked in contemporary cognitive science), and there is room for debate about what distinguishes models, specifically, from other representations. Lake et al. suggest that model-building is closely connected to anticipation, writing that, 'cognition is about using… models to understand the world, to explain what we see, to imagine what could have happened that didn't, or what could be true but isn't, and then planning actions to make it so' (2017, p. 2). This suggests a conception of models in terms of the kinds of features that they represent (i.e. their content), as opposed to the conception based on representational format which has been developed in much recent philosophical work (e.g. Swoyer 1991, Ryder 2004, Ramsey 2007, Kiefer & Hohwy 2018). On the face of it, action for reasons as model-based RL offers a straightforward example of the anticipatory conception: the thought would be that to act for a reason involves anticipating that one's action will bring about a desired result (perhaps because of some specific feature of the current environment), and that this is reflected in the point that model-based RL learns and employs representations of action-outcome contingencies.

However, two complications arise. First, not all reasons for action concern consequences. For example, I might perform a ritual for the reason that in doing so I will honour one of my deceased ancestors. In this case, there seems to be no need for me to anticipate that any particular outcome will follow from my action in order for it to have been done for a reason. To accommodate this point, careful philosophers have formulated the Humean Theory of Motivation so as to allow that actions done for reasons may be motivated either by beliefs about the likely consequences of those actions, or beliefs about what the actions may constitute. Performing actions like my ritual still seems to require something like model-based RL, because it relies on the capacity to combine beliefs about one's actions with desires, but this point does show that anticipation of consequences is not what distinguishes acting for a reason from other forms of action.

Second, the claim that representing features of the environment that facilitate anticipation is what distinguishes cognitive models is also doubtful, at least in the case of the distinction between model-based and model-free RL. In model-free RL, the agent learns and represents how much reward is likely to follow from performing actions under given conditions. This information is used both in selecting actions, and in the learning process (in which the actual level of reward received on a particular occasion is compared with the level expected). Again, this point does not call into question the apparent connection between model-based RL and acting for reasons, because one's reason for acting in a particular way cannot be that one expects it to be rewarding. Reasons for action must have the potential to justify those actions, and since 'rewarding' in this context means nothing more than 'choiceworthy', such a justification would be trivial. It also does not mean that we cannot meaningfully distinguish between model-based and model-free RL. But it does mean that a more careful account is required of the role of anticipation in distinguishing those algorithms that employ models.

References

Craik, K. 1943. *The Nature of Explanation*.

Daw, N., Y. Niv & P. Dayan. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience* 8: 1704-1711.

Dretske, F. 1993. Can intelligence be artificial? *Philosophical Studies* 71: 201-216.

Kiefer, A. & J. Hohwy. 2018. Content and misrepresentation in hierarchical generative models. *Synthese* 195 (6): 2387-2415.

Lake, B., T. Ullman, J. Tenenbaum & S. Gershman. 2017. Building machines that learn and think like people. *Behavioral & Brain Sciences* 40: e253.

Ramsey, W. 2007. *Representation Reconsidered*.

Rosen, R. 1985. *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations*.

Ryder, D. 2004. SINBAD neurosemantics: A theory of mental representation. *Mind & Language* 19: 211-240.

Smith, M. 1987. The Humean theory of motivation. *Mind* 96: 36-61.

Swoyer, C. 1991. Structural representation and surrogative reasoning. *Synthese* 87: 449-508.