The Elusive Promise of AI
By
Dr. Jeff Riley

Since the phrase *artificial intelligence* was coined in 1956, expectations of smart machines have been high. The public at large wasn't just interested in faster and smaller computers – they were expecting AI to deliver on its (largely unspecified) promise of machines that think; machines that display human intelligence. The expectation continues, but the reality is that it will probably be a very long time before AI delivers to the expectation.

In some recent work I studied the ability of machines to learn complex tasks, and as part of that work undertook an analysis of the difficulty some problems pose for reward-based learning strategies – in particular evolutionary algorithms. One of the outcomes of that study is the notion of what I have termed the "evolutionary quantum", and the conjecture that the amount of learning an evolutionary algorithm is capable of in any reasonable timeframe is inherently limited.

The problem I chose to study was that of trying to teach a simulated robot soccer player to kick a goal – *from scratch*. The method I used was to have a messy-coded genetic algorithm evolve a set of fuzzy rules that defined the behaviour of the robot soccer player. Learning to solve a problem from scratch is extremely difficult. In the case of training a robot soccer player to score goals, at the start of the learning process and without any prior expert knowledge, players tend to either do nothing or move about randomly. Since the learning algorithm is reward-based no learning can take place until a reward is earned, and if the player is only rewarded when a goal is scored it takes a very long time and a large amount of luck for any progress to be made.

In fact the results of my work indicate that there is a limit to what can be learned from any given starting point (the *evolutionary quantum* for evolutionary algorithms), and it is only when expert human knowledge is added that the algorithm can learn to solve anything but small problems in any reasonable timeframe. For evolutionary algorithms, expert knowledge is often provided in the form of some innate knowledge (e.g. a robot soccer player already knowing how to kick the ball, or that in order to score a kick needs to be aimed at the goal), or by effectively guiding learning by using an incremental reward function (e.g. rewarding a robot soccer player a small amount for kicking the ball, more for kicking the ball more than once, more still for kicking the ball close to the goal, and most for scoring a goal).

Without giving a robot soccer player innate skills (finding the ball, intercepting the ball, kicking the ball towards the goal, etc.), or guiding the player's learning by implementing incremental rewards, it is virtually impossible to teach the player to score a goal. There is a lot of expert knowledge and experience bound up in the innate skills and incremental reward function - not only is there the knowledge of how to perform the skills or actions (kicking, intercepting etc.), but there is inherent knowledge that those skills are even beneficial for producing goal-scoring behaviour. Giving the player innate skills and implementing an incremental reward function is effectively breaking the very difficult problem of learning goal-scoring behaviour into smaller, more easily solved parts, and then providing a recipe to solve those smaller parts. The real intelligence is in knowing that the problem needs to be broken into smaller parts, determining what those parts are, and providing an incremental path to solving them. Once that is done, the problem is almost guaranteed to be solved – but in doing so we lose the *artificial* part of artificial intelligence. The robot soccer problem is just one specific example, but I believe it does highlight a problem with many of the AI techniques we have today – they require human experience and know-how.

With the ever-increasing speed and computing power of modern computers we may be able to construct smart machines for specific problems (e.g. autonomous vehicle control, credit card fraud detection, etc.), and to be sure the complexity of the problems for which smart machines are deployed is increasing as we progress, but will we ever construct machines that can learn for themselves *from scratch* – machines that can truly reason? Can AI deliver what has so far proven to be a very elusive promise? I think the answer is yes – the only impediment being time. After all humans exist, so evolution has already solved the problem.