ARTIFICIAL INTELLIGENCE

Learning to see and act

An artificial-intelligence system uses machine learning from massive training sets to teach itself to play 49 classic computer games, demonstrating that it can adapt to a variety of tasks. SEE LETTER P.529

BERNHARD SCHÖLKOPF

Improvements in our ability to process large amounts of data have led to progress in many areas of science, not least artificial intelligence (AI). With advances in machine learning has come the development of machines that can learn intelligent behaviour directly from data, rather than being explicitly programmed to exhibit such behaviour. For instance, the advent of 'big data' has resulted in systems that can recognize objects or sounds with considerable precision. On page 529 of this issue, Mnih *et al.*[1] describe an agent that uses large data sets to teach itself how to play 49 classic Atari 2600 computer games by looking at the pixels and learning actions that increase the game score. It beat a professional games player in many instances — a remarkable example of the progress being made in AI.

In machine learning, systems are trained to infer patterns from observational data. A particularly simple type of pattern, a mapping between input and output, can be learnt through a process called supervised learning. A supervised-learning system is given training data consisting of example inputs and the corresponding outputs, and comes up with a model to explain those data (a process called function approximation). It does this by choosing from a class of model specified by the system's designer. Designing this class is an art: its size and complexity should reflect the amount of training data available, and its content should reflect 'prior knowledge' that the designer of the system considers useful for the problem at hand. If all this is done well, the inferred model will then apply not only for the training set, but also for other data that adhere to the same underlying pattern.

The rapid growth of data sets means that machine learning can now use complex model classes and tackle highly non-trivial inference problems. Such problems are usually characterized by several factors: the data are multidimensional; the underlying pattern is complex (for instance, it might be nonlinear or changeable); and the designer has only weak prior knowledge about the problem — in particular, a mechanistic understanding is lacking.

The human brain repeatedly solves non-trivial inference problems as we go about our daily lives, interpreting high-dimensional sensory data to determine how best to control all the muscles of the body. Simple supervised learning is clearly not the whole story, because we often learn without a 'supervisor' telling us the outputs of a hypothetical input–output function. Here, 'reinforcement' has a central role in learning behaviours from weaker supervision. Machine learning adopted this idea to develop reinforcement-learning algorithms, in which supervision takes the form of a numerical reward signal[2], and the goal is for the system to learn a policy that, given the current state, determines which action to pick to maximize an accumulated future reward.

Mnih *et al.* use a form of reinforcement learning known as Q-learning[3] to teach systems to play a set of 49 vintage video games, learning how to increase the game score as a numerical reward. In Q-learning, $Q^*(s,a)$ represents the accumulated future reward, Q^*, if in state s the system first performs action a, and subsequently follows an optimal policy. The system tries to approximate Q^* by using an artificial neural network — a function approximator loosely inspired by biological neural networks — called a deep Q-network (DQN). The DQN's input (the pixels from four consecutive game screens) is processed by connected 'hidden' layers of computations, which extract more and more specialized visual features to help approximate the complex

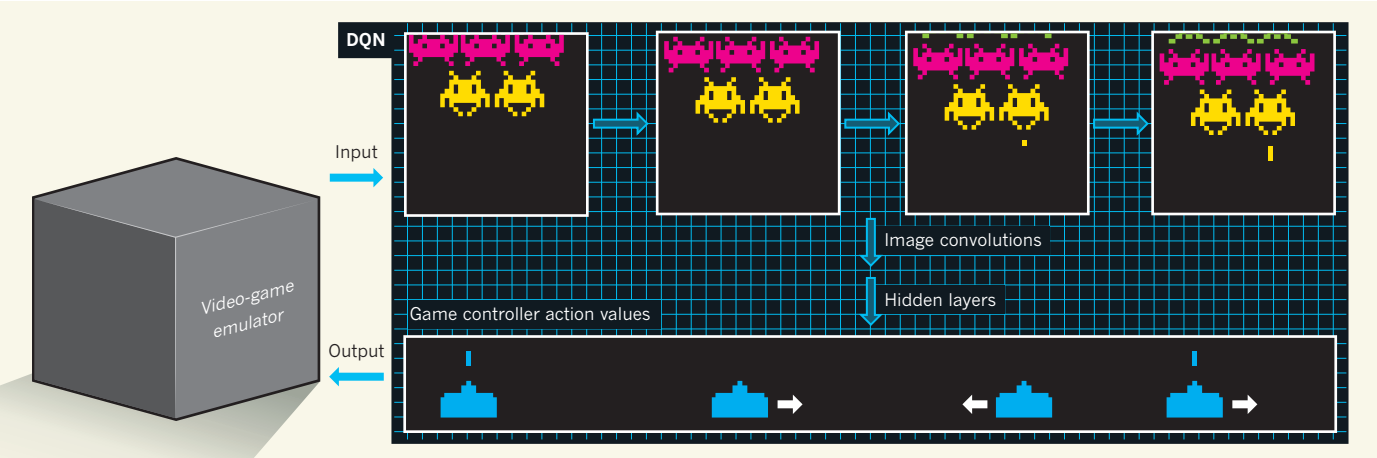

Figure 1 | Computer gamer. Mnih *et al.*[1] have designed an artificial-intelligence system, using a 'deep Q-network' (DQN), that learns how to play 49 video games. The DQN analyses a sequence of four game screens simultaneously and approximates, for each possible action it can make, the consequences on the future game score if that action is taken and followed by the best possible course of subsequent actions. The first layers of the DQN analyse the pixels of the game screen and extract information from more and more specialized visual features (image convolutions). Subsequent, fully connected hidden layers predict the value of actions from these features. The last layer is the output — the action taken by the DQN. The possible outputs depend on the specific game the system is playing; everything else is the same in each of the 49 games.

nonlinear mapping between inputs and the value of possible actions — for instance, the value of a move in each possible direction when playing *Space Invaders* (Fig. 1).

The system picks output actions on the basis of its current estimate of Q^*, thereby exploiting its knowledge of a game's reward structure, and intersperses the predicted best action with random actions to explore uncharted territory. The game then responds with the next game screen and a reward signal equal to the change in the game score. Periodically, the network uses inputs and rewards to update the DQN parameters, attempting to move closer to Q^*. Much thought went into how exactly to do this, given that the agent collects its own training data over time. As such, the data are not independent from a statistical point of view, implying that most of statistical theory does not apply. The authors store past experiences in the system's memory and subsequently re-train on them — a procedure they liken to hippocampal processes during sleep. They also report that the system benefits from randomly permuting these experiences.

There are several interesting aspects of Mnih and colleagues' paper. First, the system performances are comparable to those of a human games tester. Second, the approach displays impressive adaptability. Although each system was trained using data from one game, the prior knowledge that went into the system design was essentially the same for all 49 games; the systems essentially differed only in the data they had been trained on. Finally, the main methods used have been around for several decades, making Mnih and colleagues' engineering feat all the more commendable.

What is responsible for the impressive performance of Mnih and colleagues' system, also reported for another DQN[4]? It may be largely down to improved function approximation using deep networks. Even though the size of the game screens produced by the emulator is reduced by the system to 84×84 pixels, the problem's dimensionality is much higher than that of most previous applications of reinforcement learning. Also, Q^* is highly nonlinear, which calls for a rich nonlinear function class to be used as an approximator. This type of approximation can be accurately made only using huge data sets (which the game emulator can produce), state-of-the-art function learning and considerable computing power.

Some fundamental issues remain open, however. Can we mathematically understand reinforcement learning from dependent data, and develop algorithms that provably work? Is it sufficient to learn statistical associations, or do we need to take into account the underlying causal structure, describing, say, which pixels causally influence others? This may help in finding relevant parts of the state space (for example, identifying which sets of pixels form a relevant entity, such as an alien in *Space Invaders*); in avoiding 'superstitious' behaviour, in which statistical associations may be misinterpreted as causal; and in making systems more robust with respect to data-set shifts, such as changes in the behaviours or visual appearance of game characters[3,5,6]. And how should we handle latent learning — the fact that biological systems also learn when no rewards are present? Could this help us to handle cases in which the dimensionality is even higher and the key quantities are hidden in a sea of irrelevant information?

In the early days of AI, beating a professional chess player was held by some to be the gold standard. This has now been achieved, and the target has shifted as we have grown to understand that other problems are much harder for computers, in particular problems involving high dimensionalities and noisy inputs. These are real-world problems, at which biological perception–action systems excel and machine learning outperforms conventional engineering methods. Mnih and colleagues may have chosen the right tools for this job, and a set of video games may be a better model of the real world than chess, at least as far as AI is concerned. ■

Bernhard Schölkopf *is at the Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany.*
e-mail: bs@tuebingen.mpg.de

1. Mnih, V. *et al. Nature* **518,** 529–533 (2015).
2. Sutton R. S. & Barto A. G. *Reinforcement Learning: An Introduction* (MIT Press, 1998).
3. Watkins, C. J. C. H. *Learning from Delayed Rewards.* PhD thesis, Univ. Cambridge (1989).
4. Guo, X., Singh, S., Lee, H., Lewis, R. L. & Wang, X. *Adv. Neural Inf. Process. Syst.* 27 (2014).
5. Bareinboim, E. & Pearl, J. in *Proc. 25th AAAI Conf. on Artificial Intelligence* 100–108 (2011).
6. Schölkopf, B. *et al.* in *Proc. 29th Int. Conf. on Machine Learning* 1255–1262 (Omnipress, 2012).

BIODIVERSITY

The benefits of traditional knowledge

A study of two Balkan ethnic groups living in close proximity finds that traditional knowledge about local plant resources helps communities to cope with periods of famine, and can promote the conservation of biodiversity.

MANUEL PARDO-DE-SANTAYANA & MANUEL J. MACÍA

Understanding how human groups obtain, manage and perceive their local resources — particularly the plants they use as food and medicine — is crucial for ensuring that those communities can continue to live and benefit from their local ecosystems in a sustainable way. The study of these complex interactions between plants and people is the aim of an integrative discipline known as ethnobotany, which is based on methods derived mainly from botany and anthropology[1]. Most ethnobotanical research reveals that traditional knowledge about local edible and healing resources is suffering an alarming decline[2], especially in Europe[3]. However, writing in *Nature Plants*, Quave and Pieroni[4] suggest that wild plants still have an essential role for communities living in the mountains of Kukës, one of the poorest districts of Albania. Their results also show how preserving local knowledge is linked to maintaining biodiversity.

The mountains of Kukës lie in the Balkans, a hotspot of cultural and biological diversity that has suffered major political and economic shifts over the past three decades. Quave and Pieroni studied two culturally and linguistically distinct rural Islamic ethnic groups (the Gorani and Albanians) that, despite living in close proximity in this region and facing similar environmental and economic conditions, have remained relatively isolated from one another. The two groups use wild plants in different ways, giving the authors an opportunity to investigate the role of cultural factors in shaping how the local flora is understood and used in daily life, health practices and, ultimately, survival. Among the various quantitative techniques used, the authors designed a simple but innovative tool to compare the cultural similarities and differences between the two groups' use of plant species.

The researchers report significant variation in the plant species used for medicinal purposes by the two ethnic groups. A plausible explanation for this is that the spread of health-related lore requires a high degree of affinity, because trying a new remedy requires a great deal of trust[5]. Health is a sensitive topic, so people accept advice mainly from knowledgeable relatives or friends belonging to the same ethnic group[6]. Moreover, many traditional remedies have a highly symbolic component, and the mechanisms by which they are believed to bring about healing can lie — totally or partially — in the remedy's cultural meaning[7].

Quave and Pieroni find only two species,