Artificial Scientists

As a consequence of the scale and rate of data generation in science, of which some examples were outlined in Part 1, models of the data are increasingly requiring automatic construction and modification. Moreover, it is already clear that computing and computer science are set to play an increasingly central role in supporting the fundamental formulation and testing of scientific hypotheses. This traditionally human activity has already become unsustainable in the biological sciences without the aid of computers. This is not primarily due to the scale of the data involved but is because scientists are not able to conceptualise the breadth and depth of the relationships and potential relationships contained within the data.

Machine Learning systems that produce human-comprehensible hypotheses from data will increasingly be used for knowledge discovery within science. Such systems today are typically open loop, with no direct link between the machine learning system and the collection of data. A more closed-loop approach was investigated in the early 1990s in work on automating chemical experiments [27], though the approach was limited to the estimation of chemical parameters. However, recent advances in computer science go considerably beyond such approaches, pointing the way to an exciting and intelligent future – one of 'autonomous experimentation'. In the new approach, artificial intelligence techniques are employed to carry out the entire cycle of scientific experimentation, including the origination of hypotheses to explain observations, the devising of experiments to test these hypotheses and the physical implementation of the experiments using laboratory robots to falsify hypotheses. Such a system has already been demonstrated in the 'Robot Scientist' project [28] where laboratory robots conducted experiments selected by 'active learning'.

The Robot Scientist project demonstrated that using a machine learning system, based on Inductive Logic Programming (ILP), the robot selected experiments to discriminate between contending hypotheses. The experiments were based on gene knock-outs for yeast (*Saccharomyces cerevisiae*). The aim was to determine the function of the gene by varying quantities of nutrient provided. Feedback on the outcomes of experiments came in the form of indications of whether or not the yeast died within a 24-h period. The robot's intelligent experiment selection strategy based on the ASE-Progol system was competitive with human performance and significantly outperformed both cheapest and random-experiment selection with a cost decrease of 3-fold and 100-fold, respectively.

One exciting development we might expect to see in this area over the next 10-years is the construction of the first micro-fluidic Robot Scientist. This would involve the confluence of active learning and autonomous experimentation systems with micro-fluidic technology (see section on 'Molecular Machines' below). The key effects of miniaturising the technology would be the reduction of the experimental cycle time from hours, in the case of the previously published Robot Scientist, to milliseconds, with a corresponding increase in the robustness of outcomes from micro-fluidic experiments. Further flexibility could be added to such a scenario by the use of Chemical Turing machines (see next section) to allow automatic online preparation of a wide variety of chemical compounds as input to each experiment.

Conducting impossible experiments

Autonomous experimentation will undoubtedly play an important role towards 2020 in meeting the challenge of accumulating and analysing the comprehensive data sets (e.g. such as is required for biological modelling at the system level) that are outside 'normal' affordable resource constraints including time and human bandwidth.

Moreover, such computational techniques capable of deciding from past observations which test to perform next are also of great interest for instrumentation with limited communication bandwidth where decisions need to be taken remotely as to which is the best 'next step' to take, such as which experiment to perform, when to perform it or when to stop collecting data. Mobile robots exploring remote or harsh environments such as deep sea locations or other planets typically are able to communicate back only a small fraction of the data gathered by their sensors. Accordingly, the robot itself would be in the best position to decide the next experimental step.

In the coming decade, we can expect a confluence of wireless networks and lab-on-chip sensor technology with large-scale applications in environmental monitoring. Autonomous experimentation will be required to fully exploit the potential of this technology. In such a lab-on-chip network, each sensor node is endowed with a limited supply of wet chemistry and accordingly can perform only a limited number of experiments. The network will collectively decide how these resources should be spent.

In summary, autonomous experimentation enabled by intelligent 'Artificial scientists' has begun to open up highly novel experimental approaches that have previously been unimaginable.

Stephen Muggleton, Klaus-Peter Zauner