Could a Computer Create its Own Representations?

Could a Computer Create its Own Representations?

Alan Bundy
School of Informatics
University of Edinburgh

Artificial Intelligence systems need to maintain a representation of their environment so that they can interpret sensory information and plan actions. This is true for robots, chess playing programs, question-answering systems, softbots on the web, in fact, any computer-based agent. Most such representations are initially hand crafted by the system's developer then incrementally adapted by the addition and deletion of facts, as sensors detect environmental changes or actions make such changes.

This arrangement was sufficient when agents were built only for a narrowly defined task in a fairly stable environment, but it is no longer sufficient for many applications.

- In a rapidly changing environment, it will not be sufficient to add and delete a few facts in a representation. Not only must rules also evolve, but change to the representation's language may be needed. In a multi-agent environment, agents must represent each other. Not only might agents change, but so might the agent population.
- Problem solving success is very dependent on problem representation. If an agent's tasks change then its old representation may no longer suit its new tasks, so it must evolve to suit the new problems.

For example, suppose you have a softbot agent on the web that makes plans for you using the services offered by other agents. It might plan a holiday by combining taxi, plane, accommodation, and sightseeing services. The population of service-providing agents is huge and rapidly changing. The kind of tasks you set your agent may change. Your agent must be able to evolve its representation to match this changing environment.

My research group has been developing algorithms to enable agents to change their representations. We are especially interested in conceptual changes driven by reasoning failures. For instance, reasoning may fail because it infers something false, fails to conclude something true or is just very slow at concluding anything. Diagnosis of such failures can suggest repairs to the agent's faulty representation. The change required might just be to add or delete a fact, but it might instead be to provide a missing precondition to a rule or to change the *language* of the representation.

In our softbot world, reasoning failures take the form of plans that fail on execution. The planning agent may then need to delete the 'fact' that another agent offers a service, or it may need to add a new precondition that another agent requires to be paid in advance. More fundamentally, agents may fail to communicate because they do not share a vocabulary. Their vocabularies must then be aligned.

We have developed algorithms for automating representational change in the following domains: web-based, service-providing agents; discrepancies between physics theories and experimental evidence; and mathematical proofs. Recently, we have generalised from these domain-specific algorithms to develop the general-purpose *reformation* algorithm.

In all these applications, the following simple language changes were sufficient.

- One concept is split into several, or several concepts are merged into one. Examples from physics are: splitting ‘matter’ into ‘dark matter’ and ‘visible matter’; and merging of ‘morning star’ and ‘evening star’ into ‘Venus’.
- A dependency may need to be added or deleted. For example, the period of a pendulum *does* depend on its length, but the acceleration of an object in free fall *does not* depend on its weight.

Such changes may appear minor but, as the history of physics shows, they can accumulate into significant changes.

Recently, we have explored algorithms for *analogical blending*, where two old concepts are merged into a new one. For instance, we can merge ‘house’ and ‘boat’ to form either ‘houseboat’ or ‘boathouse’. ‘Houseboat’ comes from aligning the house with the boat, but ‘boathouse’ comes from aligning the boat with the house’s occupant.

Analogical blends can be used to form novel concepts; an inadequate representation may be improved by boosting it with concepts from an analogous representation. An example, again from physics, is Rutherford’s model of the atom as a miniature solar system, in which the nucleus is aligned with the Sun and the electrons with the planets. Often, analogical blends are faulty and need repairing. This was true of Rutherford’s model, which failed to explain why the electrons did not emit radiation, lose energy and fall into the nucleus. The model was repaired with the aid of quantum mechanics. We have recently begun exploring how reformation can be used to repair faulty blends.

Can our automated agents be claimed to create their own representations? All our algorithms only evolve old representations into new ones. There is no *de novo* creation of representations. But would it be reasonable to expect one? Human new representations also seem to come by analogy with, or repair of, old representations, informed by their failures.