

Towards an Adaptive Intelligent Agent

Daniel Oblinger

Information Processing Techniques Office, Defense Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714, USA
Daniel.Oblinger@darpa.mil

We consider the learning required for existing intelligent biological agents (human and animal), and partition the learning required to produce such an agent into: (i) genomic, (ii) cultural, and (iii) individual learning. Using this simple taxonomy we consider limits inherent on the complexity of genomic knowledge, and the fact that many complex tasks did not exist during evolutionary times. Both suggest the necessity of some form of universal bootstrapping to form the basis of post-genomic (cultural and individual) learning.

Two common families of approaches to learning agents are either to use weak universal learning (e.g. Reinforcement learning) or strong task-specialized learning and reasoning systems (e.g. case based or explanation based reasoning systems). Here we consider a third path, one that is universal, in the sense of weak learning systems, but take advantage of several reinforcing feedback signals that permit bootstrapping of strong performance from repeated iterations of weak learning.

While not a common approach, “growing” an intelligent agent – the notion of iteratively using a strong and universal bootstrapping type algorithms – certainly have been considered in one form or another. Here, rather than try to completely specify a proposed system that fits this framework, we try to analyze what we know today about the properties that such a system must possess. We consider growing evidence from the field of neuroscience on bounds on the complexity of the genomic learning implicit in the organization of the biological agents. We then turn to consider what tasks one can expect such a universal bootstrapping algorithm should be able to solve. After all, if this proposed approach is correct, then all biological agents are examples of such bootstrapping systems, and we can reason about which tasks solvable by biological agents today were available to these agents during evolutionary times.

Given constraints from both of these directions, we consider possible architectures that could serve to bootstrap an intelligent agent with our order of complexity. As said above, rather than proposing a fully specified system, we strive to capture the important sources of

constraint that would drive such systems. It is our belief that many instantiations will ultimately be found that can bootstrap intelligence, but all will leverage these key sources of constraint. The first key source used by these systems is the use of a perceive/abstract/act loop to allow an agent to supervise its own learning. Embedding learning in this way affords a system the strength of supervised learning, that is, feedback on the correctness of its generalizations, but in a way that does not use a teacher for the labeling, but rather allows the agent to use its interaction with the world (obtained in an unsupervised fashion) to drive this labeling for stronger learning. Second, several key “emotional” drivers can serve to structure the learning of an unsupervised agent in a way that affords it feedback signals which facilitate indefinite bootstrapping. Third, a class of social “emotional” drivers serves to structure the interactions between agents in order to drive cultural learning in a way that facilitates the transfer of cultural knowledge from one agent to another.

Taken together, we argue, it is at least plausible that a relatively simple collection of learning processes and adaptive driving feedback processes could bootstrap from this seed to a complex adapted intelligent agent.