Bootstrapping Communications from Shared Experience

Jake Beal

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Consider two agents connected by a large bundle of wires with an unknown permutation. The two agents wish to communicate and do not have a language with which to do so, but observe similar events in their external environments. I am developing an algorithm which allows them to rapidly develop a simple language, based on the assumption that their external observations are correlated.

Motivation: Neuroscience has postulated that the brain has many "organs" — internal subdivisions which specialize in one area. If we accept this view, then we need some sort of mechanism to interface these components. The design of this mechanism is limited by the hardware which the brain is constructed out of, as well as the size of the blueprints specifying how it is built. Neurons, as hardware, are relatively slow and imprecise devices, but they are very cheap, and it's easy to throw a lot of them at a problem in parallel. Our DNA is only about 1 gigabyte, too small to encode the full complexity of interfaces between all of the different components.

This problem, then, assumes we have two regions of the brain connected by a twisted nerve bundle. An algorithm solving this problem will give us insight and a tool for thinking about how a brain might be composed of independent parts which learn to cooperate with each other.

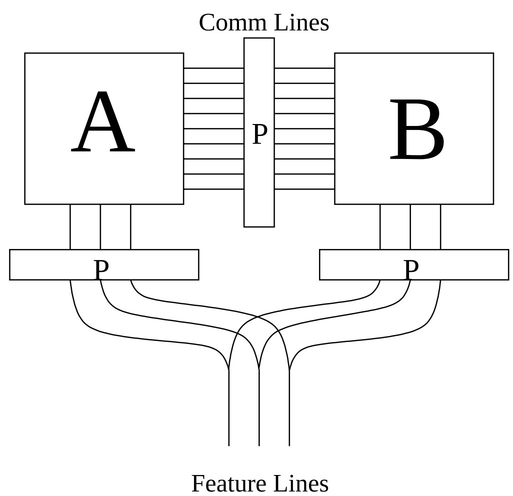

Figure 1: The agents labelled A and B are interconnected by *comm lines* — a bundle of wires with an arbitrary and unknown permutation. The agents also share some *feature lines* with the outside world, again with unknown permutations.

Previous Work: This work is inspired in part by Simon Kirby's experiments on language evolution. In these experiments, Kirby showed that, a population of grammar induction agents, given shared semantic information, could evolve a simple language from initial noise.[1]

Another key inspiration for this work is Yip & Sussman's work on one-shot learning via sparse representations, which lays out a program of building hardware components which embody some piece of "intelligent" behavior. These components, once developed, may then form a "TTL Databook of the Mind" which we could use to build more complex and intelligent structures with relative ease.[2]

Approach: I am attacking this problem by taking advantage of the large number of wires connecting the two agents. Any symbol made up by an agent uses only a few of these wires, selected randomly. The sparseness of these symbols then allows stimuli to be separated easily by a listening agent by intersection of multiple transmissions. Transmission of symbols is then upgraded to transmission of thematic-role frames by inflecting the symbols to indicate the role played by a symbol.

This approach allows the agents to develop a shared language of s symbols and i inflections in $O(s+i)$ time. Moreover, the learning time per symbol is very fast - on the order of a dozen exposures to a given symbol. Since, in general, multiple symbols are seen at once, this leads to an extremely rapid convergence time for the algorithm.

Impact: The ability for two agents to bootstrap a communications system so quickly from shared observations only gives insight into how the human brain might be organized, and provides powerful tools for thinking about the problems of building intelligent machines of our own design.

Future Work: At present, I have a functioning prototype which learns vocabulary and role-inflections in low-constant linear time. The next step is to more tightly analyze its performance and capabilities, in order to better grasp the principles behind it.

References:

[1] Simon Kirby. Language evolution without natural selection: From vocabulary to syntax in a population of learners. Technical Report EOPL-98-1 Edinburgh Occasional Paper in Linguistics, University of Edinburgh Department of Linguistics, 1998.

[2] Kenneth Yip and Gerald J. Sussman. Sparse representations for fast, one-shot learning. Technical Report 1633, MIT Artificial Intelligence Laboratory, May 1998.