TTL Databook of the Mind

Jake Beal & Gerald Jay Sussman

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: This is a long-term effort to develop components which can serve as a foundation from which complex "intelligent" systems can be built, much the way that the TTL databook served as a foundation from which complex digital electronic systems could be built. We want to have a similar library of modular components, each able to perform some well-specified aspect of "intelligent" behavior.

Motivation: Recent research suggests that the human brain may be a highly organized structure with many specialized components, rather than some sort of vast disorganized "neural pudding." The human genetic code, however, is only about 1 gigabyte, too small to encode the full structure of the brain in detail. One way of reconciling the structure of the brain with the poverty of instructions available for constructing it is to imagine the brain as composed of a large number of small components, where each component has a well-specified design drawn from a relatively small library of component types.

In building a TTL databook for the mind, we hope to attack the problem of intelligence on a median level of complexity: high-level in the sense that we are not attempting to model neural hardware, but are concerned with pieces of functionality, low-level in the sense that we are concerned with building simple mechanisms capturing only small fragments of "intelligence".

From the first simple components, we hope to build more complex components, then more complex components from those, and so on, just as flip-flops and xor gates combine to form components like adders and muxes, which combine to form components like ALUs and UARTs, from which we build computers. This is the analogy we are following for the TTL databook of the mind.

Previous Work: Major inspiration for this probably lies in Minsky's Society of Mind, which approaches the problem of human intelligence from a hardware design perspective. Minsky considers many different questions of intelligence and proposes simple mechanisms for implementing each. The TTL project extends this idea by looking for a unified framework in which the mechanisms is produces can interact.[3]

More recently, two projects have yielded components usable as part of the TTL project. First of these is the work by Yip & Sussman on one-shot learning via sparse representations. This work demonstrates a mechanism where constraints operating on a shift-register are able to learn English pluralization and past tense rules from a few examples.[5]

Second is work by Beal on building communication systems based on shared observations of the external world. This work demonstrates a mechanism where agents use low-weight signals on a large bundle of communication lines to rapidly generate a shared vocabulary in which to communicate.[1]

Approach: Mechanisms designed under the TTL program should conform to several biologically inspired constraints. These constraints are imposed to help avoid the tendencies over-generality and hidden dependences which often cripple approaches to AI problems - they are not absolute requirements, but tools to guide thinking about the problem. In short, we see that the human brain is successful, so we constrain ourselves with its limitations and benefits, to improve our chances of discovering some of the tricks which it uses.

- **Plentiful hardware:** Neurons are cheap hardware, and it is easy to throw a lot at a problem in parallel. TTL projects should not shirk from throwing massive amounts of hardware at a problem.

- **Short serial paths:** Neurons are quite slow, compared to digital computers. Nevertheless, the human brain works quite well - we should force ourselves to look for the solutions it might use, rather than taking advantage of the serial speed of computers.

- **Modeless operation:** Humans do not operate on a query-response basis, nor do they have clear "modes of operation". Accordingly, neither should the mechanisms we design.
- **Local control:** Hardware for any part of the system should operate independently, not at the behest of some central agency.
- **Limited globals:** The human brain does not appear to work in tight synchrony, nor does it have more than a few signals with wide distribution. Accordingly, our mechanisms should only use such tools if they are implemented by a distributed local system.
- **Invertible mechanisms:** Parsimony suggests that, for example, the same lexicon interpreting sets of heard phonemes as words ought to render words into phonemes being spoken. The brain also appears to show symmetry in structure suggesting that input hardware may be used for output and vice versa.
- **Gradual performance degradation:** The mechanisms we build should be able to tolerate noise and error, just as humans can. Moreover, failure should be gradual, as a slow degradation of performance, rather than a systems crash.
- **Rapid learning:** Humans can learn from a single example. Our systems should be able to as well.
- **Clear embodiment of knowledge:** This is not a biological constraint, but one allowing us to know when we have succeeded. It should be possible to examine any TTL mechanism and discover exactly what it has learned. Moreover, this data should evidence correct understanding of structure, rather than mere reproduction of behavior.
- **Non-probabilistic Behavior:** Another non-biological constraint. The system should not make choices based on probability, as this tends to obscure its knowledge rather than reveal it.

Impact: If we are successful, we expect to have a revolutionary effect on the understanding of intelligence and our capacity to engineer artificial intelligence systems.

Future Work: I am looking to investigate a mechanism for storing knowledge and making simple deductions. This work is based on a revisitation of Winston's analogy-reasoning[4] and Falhman's NETL[2], from the TTL perspective.

References:

[1] Jacob Beal. An algorithm for bootstrapping communications. Technical Report 2001-016 (in press), MIT Artificial Intelligence Laboratory, August 2001.

[2] Scott E. Fahlman. *NETL: A System for Representing and Using RealWorld Knowledge*. MIT Press, 1979.

[3] Marvin Minsky. *The Society of the Mind*. Simon & Schuster, Inc, 1985.

[4] Patrick H. Winston. Learning by augmenting rules and accumulating censors. Technical Report 678, MIT Artificial Intelligence Laboratory, May 1982.

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