Leveraging Language into Learning

Jacob Beal
MIT CSAIL
77 Massachusetts Ave
Cambridge, Massachusetts 02139
jakebeal@mit.edu

Abstract

I hypothesize that learning a vocabulary to communicate between components of a system is equivalent to general learning. Moreover, I assert that some problems of general learning, such as eliminating bad hypotheses, deepening shallow representations, and generation of near-misses, will become simpler when refactored into communication learning problems.

I hypothesize that learning and reasoning can be a byproduct of translation between different perspectives. When two agents with different perspectives learn to communicate, some of the structure of the world relating their perspectives will be encoded into the communication system they generate. If the two agents later communicate to bring their models of a situation into alignment, then the process of translation will effectively reason with the knowledge encoded in the communication system.

The imperfect translation certain in such a system is an advantage rather than a limitation. Attempts at translation may serve to check for mistakes, refactor difficult problems, enable learning of deeper structure, and even generate original ideas.

In my master's thesis, I demonstrated two agents creating a shared vocabulary and inflections to communicate thematic role frames.(Beal 2002a; 2002b) The two agents use very similar representations, so only identity relations are learned. With some generalization, other relations such as cause and effect or proximity should be nearly as easy to acquire. I propose to demonstrate that a group of agents, each responsible for one representation (e.g. vision, language, motor, social), can acquire a communication system that embodies knowledge about the world and act reasonably in complex situations as a byproduct of translating between representations.

Motivation

"If you've found the right representation, the problem is already almost solved." –Patrick Winston

Our programs are fragile and have a hard time adapting to changing situations. People, on the other hand, are very good at adapting, so I look to human intelligence for inspiration on this problem.

How is it that people can deal with confusing, unfamiliar situations? One thing they do is talk with other people! People brainstorm, free-associating ideas and trying to make sense of the detritus they produce. Teachers explain material to their students and end up understanding it better themselves. And being wrong is sometimes more important than being right: there are stories about people going to tell Marvin Minsky about an idea, which he then misunderstood as a much better idea than they'd started with!

There is some evidence that the human brain might operate on the same principle. Cognitive science research finds that the human brain may be composed of modules(Kanwisher 1998; Kanwisher & Wojciulik 2000) which learn throughout childhood to communicate and to reason cooperatively.(Spelke, Vishton, & von Hofsten 1996; Hermer & Spelke 1996) Even in adults, collaboration between regions of the brain may depend on a language faculty. For example, "jamming" language prevents an adult from using color information to reorient(Hermer-Vasquez, Spelke, & Katznelson 1999) and native language biases how vision affects reasoning about time.(Boroditsky 2001)

For the human brain, the evidence is inconclusive, but it and the example of people learning by talking together lead me to conjecture a novel engineering principle: *a system composed of components which collaborate by constructing a shared vocabulary can adapt to novel or confusing situations.*

Some of the capabilities I expect such a system might have:

- **Creative Misunderstanding:** Translating between components with dissimilar representations will tend to mutate the data being translated. Since mistranslation is systematic, however, the mutation may tend to produce different sense rather than nonsense.[1]
- **Mistake Filtering:** Components with dissimilar representations will tend to make different mistakes. A group of components comparing possible conclusions can then

[1] Kirby exploits this in his work on language evolution(Kirby 1998; 2002), in which a coincidence in the speaker's utterances becomes a pattern in the listener's interpretation.

collectively filter out most incorrect conclusions, greatly reducing the search space.

- **Abstraction Generation:** Communication by constructing a shared vocabulary means that relations between features are instantiated as vocabulary words. The presence of a vocabulary word can, itself, be a feature, allowing recursive learning of more abstract relations.[2]

If my hypothesis is correct, it will have both theoretical and practical implications. On the theoretical side, the model can be applied to cooperative reasoning among regions of the human brain and tested by cognitive experiments. Practically, it will be a powerful tool for attacking representation-intensive problems, and a major step towards producing systems which exhibit human-like intelligence.

Research Plan

1. Generalize the communication bootstrapping mechanism (Beal 2002b) for systems with many components, multiple connections per component, and only partial synchronization.

   Executing this step uncovered an unexpected subproblem: differing propagation times through a loosely synchronized system can create inconsistent transitory states which confuse the system if it tries to learn from them. In solving this problem, I discovered a heuristic which appears to be generally useful for segmenting continuous experience into discrete examples.

2. Demonstrate learning of a wide variety of relations as vocabulary words shared between a set of components; in particular difference, causality, subclass/superclass, time/space proximity measure, part/whole, containment, and object attributes. To date, I have demonstrated a restricted case of learning difference relations: a single component exposed to a simulated four-way intersection learns a set of vocabulary words that collectively describe the behavior of a stoplight.

3. Demonstrate scenarios in which reasoning is carried out as a side effect of translation between components using learned vocabulary words. To date, I have two partially developed examples: a dinner scenario, in which hearing "please pass the butter" causes a hand to move a mostly-hidden dish to the speaker, and a blocks-world scenario where an arm avoids hitting the tower it's building.

4. Formalize the conditions required for learning and the principles demonstrated by reasoning in these situations.

5. Demonstrate improved capabilities in mistake filtering, near miss generation, and abstraction generation which are made possible by these principles.

References

Beal, J. 2002a. An algorithm for bootstrapping communications. In *International Conference on Complex Systems (ICCS2002)*.

Beal, J. 2002b. Generating communications systems through shared context. Technical Report AITR 2002-002, MIT.

[2]Winston's Macbeth system(Winston 1970) demonstrates the power of reified relations in metaphoric story understanding.

Boroditsky, L. 2001. Does language shape thought? english and mandarin speakers' conceptions of time. *Cognitive Psychology* 43:1–22.

Hermer, L., and Spelke, E. 1996. Modularity and development: the case of spatial reorientation. *Cognition* 61:195–232.

Hermer-Vasquez, L.; Spelke, E.; and Katznelson, A. 1999. Sources of flexibility in human cognition: Dual-task studies of space and language. *Cognitive Psychology* 39:3–36.

Kanwisher, N., and Wojciulik, E. 2000. Visual attention: Insights from brain imaging. *Natural Reviews Neuroscience* 1:91–100.

Kanwisher, N. 1998. *The Modular Structure of Human Visual Recognition: Evidence from Functional Imaging*. Psychology Press. 199–214.

Kirby, S. 1998. Language evolution without natural selection: From vocabulary to syntax in a population of learners. Technical report, Language Evolution and Computation Research Unit, University of Edinburgh.

Kirby, S. 2002. *Learning, Bottlenecks and the Evolution of Recursive Syntax.* Cambridge University Press. chapter 6.

Spelke, E.; Vishton, P.; and von Hofsten, C. 1996. *Object perception, object-directed action, and physical knowledge in infancy*. MIT Press.

Winston, P. 1970. *Learning Structural Descriptions from Examples*. Ph.D. Dissertation, MIT.