

A Connectionist Model Of Instructed Learning

David C. Noelle

Computer Science & Engineering
University of California, San Diego
La Jolla, California 92093-0114
dnoelle@cs.ucsd.edu

The focus of this research is on how people blend knowledge gained through explicit instruction with knowledge gained through experience. The product of this work will be a cognitively plausible computational learning model which integrates instructed learning with inductive generalization from examples. The success of this model will require the attainment of both a technical and a scientific goal.

The technical goal is the design of a computational mechanism in which induction and instruction are smoothly integrated. The design of such a multistrategy learner might be implemented within a symbolic rule-based framework (Huffman, Miller, & Laird 1993), within a framework strong in inductive generalization, such as connectionism (Noelle & Cottrell 1995), or within a hybrid architecture (Maclin & Shavlik 1994). This work pursues the second of these three general approaches. Following an intuition concerning the primacy of induction (which precedes linguistic rule following both phylogenetically and ontogenetically) and with an eye on future reduction to neurological explanations, the model proposed here is built within a wholly connectionist framework.

The scientific goal of this research is to account for certain interaction effects between instruction and induction that have been observed in humans, including:

- the performance improvements brought by experience following direct instruction.
- the apparent superiority of providing direct instruction *prior* to the presentation of examples.
- the relative insensitivity to contingencies exhibited by subjects when following a simple rule.
- the intrusion of similarity into instructed category learning, resulting in a failure to follow instructions.

If successful, our computational model will exhibit all of these phenomena, and it will produce testable hypotheses concerning the accuracy and time course of human learning in various contexts.

Typical connectionist models learn via the modification of connection weights in response to an external error signal. Such methods, however, are essentially too slow to account for the rapid effects of “learning

by being told”. Instead, our approach captures the effects of direct instruction in the dynamic activation state of a recurrent network. The combinatoric space of valid instruction sequences is represented by the space of *articulated attractors* in the network’s dynamics. Each fixed-point attractor in this structured activation space encodes a unique collection of instructed knowledge. Linguistic advice is viewed as input activity which *pushes* the network into an appropriate basin of attraction, allowing the network to settle to a representation of the received knowledge. By modeling explicit instruction in this way, connection weight modification is left in the capable hands of standard inductive learning techniques. This allows induction and instruction to operate in tandem, and it opens the door to complex interactions between them.

We have already demonstrated the ability of our networks to learn a simple instructional language in the service of a task and to thereafter exhibit rapid instruction following behavior. We have also begun applying our model to experimental results from the category learning and implicit/explicit learning literature. A detailed comparison of our model with some symbolic models of instructed learning, such as Instructo-SOAR (Huffman, Miller, & Laird 1993), will also be conducted.

References

- Huffman, S. B.; Miller, C. S.; and Laird, J. E. 1993. Learning from instruction: A knowledge-level capability within a unified theory of cognition. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, 114–119. Boulder: Lawrence Erlbaum.
- Maclin, R., and Shavlik, J. W. 1994. Incorporating advice into agents that learn from reinforcements. In *Proceedings of the 12th National Conference on Artificial Intelligence*, 694–699. Seattle: AAAI Press.
- Noelle, D. C., and Cottrell, G. W. 1995. A connectionist model of instruction following. In Moore, J. D., and Lehman, J. F., eds., *Proceedings of the 17th Annual Conference of the Cognitive Science Society*, 369–374. Pittsburgh: Lawrence Erlbaum.