

# Integrating Induction & Instruction: Connectionist Advice Taking

David C. Noelle and Garrison W. Cottrell

Department of Computer Science and Engineering

University of California, San Diego

La Jolla, CA 92093-0114

dnoelle@cs.ucsd.edu and gary@cs.ucsd.edu

Humans improve their performance by means of a variety of learning strategies, including both gradual statistical induction from experience and rapid incorporation of advice. In many learning environments, these strategies may interact in complementary ways. The focus of this work is on cognitively plausible models of multistrategy learning involving the integration of inductive generalization and learning "by being told". Such models might be developed by starting with an architecture for which advice taking is relatively easy, such as one based upon a sentential knowledge representation, and subsequently adding some form of inductive learning mechanism. Alternatively, such models might be grounded in a statistical learning framework appropriately extended to operationalize instruction. This latter approach is taken here. Specifically, connectionist back-propagation networks (Rumelhart, McClelland, & the PDP Research Group 1986) are made to instantaneously modify their behavior in response to quasi-linguistic advice.

Many of the previous approaches to the instruction of connectionist networks have involved the encoding of symbolic rules as initial connection weights which may be later refined by inductive learning (Giles & Omlin 1993) (Tresp, Hollatz, & Ahmad 1993). A major drawback of this approach is that advice may only be given *before* inductive training begins. This is an unreasonable constraint for a cognitive model of instructed learning. Instead, a connectionist network is needed which may have its behavior altered by a stream of encoded instructions without a delay period for lengthy retraining.

The approach which is examined here focuses on encoding the receipt of instruction as motion in a network's activation space. In short, advice is presented to such an instructable network as a temporal sequence of instruction tokens, where each token is encoded as an input activation pattern. The network is trained to appropriately modulate its behavior based on input of such advice sequences. The correct interpretation and operationalization of input instruction sequences is learned inductively, but, once this initial learning is complete, instruction following proceeds at

the speed of activation propagation. This focus on activation space dynamics allows instructional learning and standard connectionist inductive learning to function in tandem.

This strategy has been successfully applied to a simple discrete mapping task and to the learning of natural number arithmetic. In this latter domain, the connectionist adder of Cottrell and Tsung (Cottrell & Tsung 1993), which is capable of systematically operating on arbitrarily large natural numbers, was augmented to receive instruction in various methods of addition and subtraction. The resulting network tackles arithmetic problems by examining one column of digits at a time and sequentially performing actions such as writing a resultant digit for the column, announcing a carry or borrow, and shifting attention to the next digit column. The network's behavior is determined by the most recently presented sequence of instruction tokens. Future experiments will extend these multistrategy learners to include auto-associative memories containing *articulated attractors* in activation space which will facilitate systematic generalization to novel advice sequences. These later experiments will abandon arithmetic and will focus instead on simple planning tasks in a "blocks world" environment.

## References

- Cottrell, G. W., and Tsung, F.-S. 1993. Learning simple arithmetic procedures. *Connection Science* 5(1):37–58.
- Giles, C. L., and Omlin, C. W. 1993. Rule refinement with recurrent neural networks. In *1993 IEEE International Conference on Neural Networks*, 801–806. San Francisco: IEEE Neural Networks Council.
- Rumelhart, D. E.; McClelland, J. L.; and the PDP Research Group. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, volume 1. Cambridge: The MIT Press.
- Tresp, V.; Hollatz, J.; and Ahmad, S. 1993. Network structuring and training using rule-based knowledge. In Hanson, S. J.; Cowan, J. D.; and Giles, C. L., eds., *Advances in Neural Information Processing Systems 5*. San Mateo: Morgan Kaufmann.